University of Alberta

Expression, Purification and Isotopic Labelling of HAV and HRV14 3C Protease

by

Ping Li ©

A thesis submitted to the Faculty of Graduate Studies and Research in Partial fulfillment of the requirements for the degree of

Master of Science

in

Pharmaceutical Sciences

Faculty of
Pharmacy and Pharmaceutical of Sciences

Edmonton, Alberta
Fall, 2000

National Library of Canada
Acquisitions and Bibliographic Services
395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale du Canada
Acquisitions et services bibliographiques
395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-59833-0

Canada

University of Alberta
Library Release Form

Name of Author: Ping Li

Title of Thesis: Expression, Purification and Isotopic Labelling of HAV and HRV14 3C Protease

Degree: Master of Science

Year this Degree Granted: 2000

Permission is hereby granted to the University of Alberta Library to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves all other publication and other rights in association with the copyright in the thesis, and except as hereinbefore provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatever without the author's prior written permission.

P Li

Faculty of Pharmacy and
Pharmaceutical Sciences
University of Alberta
Edmonton, Alberta
T6G 2N8

Date: Sep 27, 2000

University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research of acceptance, a thesis entitled "Expression, Purification and Isotopic Labelling of HAV and HRV14 3C Proteases" submitted by Ping Li in partial fulfillment of the requirements for the degree of Master of Science in Pharmaceutical Sciences.

David Wishart

Dr. David S. Wishart

Bernard Lemire

Dr. Bernard Lemire

Dr. Len I. Wiebe

Dr. Mavanur R. Suresh

Date Sep 26, 2000

Abstract

Acute hepatitis is caused by the hepatitis A virus (HAV). The common cold is caused by related picornavirus called the human rhinovirus (HRV). Interestingly, both HAV and HRV require a virally encoded 3C protease to complete their maturation process. Because of their importance in viral maturation, it is widely believed that the HAV and HRV 3C proteases would be ideal antiviral drug targets. This thesis describes our efforts to express, label and purify milligram quantities of the HAV and HRV 3C proteases for the purpose of NMR studies. Chapter 1 provides an overview of both HAV and HRV along with a brief description of various isotopic labelling techniques. Chapter 2 describes the strategies used to specifically label HAV 3C protease with ^{15}N phenylalanine utilizing a specially developed auxotrophic *E. coli* strain. Chapter 3 details the strategies employed to clone, express and purify HRV14 3C protease in two expression vectors (pET3a and pTrcHis2). This chapter also describes a cell adaptation process designed to permit the production of HRV14 3C protease in minimal media for isotopic labelling.

Acknowledgments

First and foremost, I would like to express my sincere gratitude to my supervisor, Dr. David S. Wishart, for his guidance, generosity, and help. Without his patience and support, I would not have been able to accomplish this work. His broad range of knowledge and acute insight into biology problems have been indispensable to both my project and my thesis accomplishments.

I would also wish to express my whole-hearted thanks to my colleagues and fellow students with whom I have had the pleasure of associating with over the past two and half years. In particular, I would like to thank Sudeepa Bhattacharyya and Dr. Dong Luo for constructing the pTrcHis2-HRV14 3C protease expression system. I would also like to thank Ashenafi Abera, Hassan Monzavi, Haiyan Zhang, Alan Gibbs, Gary Van Domselaar, Lena Andrew and Trent Bjorndahl for their help.

I am grateful to Dr. Ken Ng for supplying the pET3a-HRV14 3C protease expression system, to Dr. David Waugh who supplied the auxotrophic *E.coli* stain CT16 with which we initiated the specific labelling of HAV 3C protease. I would also like to express my appreciation to members of my advisory committee, Dr. Mavanur Suresh and Dr. Len Wiebe for their friendship, advice and assistance throughout my studies. I am also grateful to Mr. Glen Zhang for his insightful comments on the HRV14 3C project.

I particularly wish to express my deep gratitude to my parents for their love, sacrifice and encouragement over the past two and half years.

Finally, I would like to acknowledge the F. S. Chia Scholarship endowed and the PENCE for financial support throughout my graduate career.

Table of Contents

	Page
Chapter 1: Introduction	1
1.1. Picornaviruses	3
1.2. Picornaviral 3C Protease	6
1.3. Structure-assisted Design of 3C Protease Inhibitors by X-ray	14
1.4. Structure-assisted Design of 3C Protease Inhibitors by NMR	18
Chapter 2: Residue Specific ^{15}N Labelling of HAV 3C Protease	
2.1. Introduction	26
2.2. Materials and Methods	28
2.3. Results	42
2.4. Discussion	43
2.5. Conclusion	51
Chapter 3: Purification and Isotopic Labelling of HRV 3C	52
3.1. Introduction	
3.2. Materials and Methods	61
3.3. Results	68
3.4. Discussion	77
3.5. Conclusion	85
Chapter 4: Conclusion	86
References:	88

List of Tables

Table No.		Page
Table 2.1.	Amino acid enriched M9 minimal media.....	29
Table 2.2.	Terrific Broth recipe.....	30
Table 2.3.	SOC transformation medium recipe.....	31
Table 2.4.	Reagents for SDS-PAGE gel.....	41
Table 3.1.	Purification profile of HRV14 3C protease.....	74

List of Figures

Figure No.		Page
Figure 1.1.	Polyprotein cleavage scheme.....	8
Figure 1.2.	HRV14 3C protease structure	10
Figure 1.3.	Active site residues of HAV 3C protease.....	13
Figure 1.4.	HRV 3C inhibitor AG7088	16
Figure 1.5.	HAV 3C inhibitors	17
Figure 2.1.	pHAV-3CEX map	33
Figure 2.2.	HAV 3C protease expression and purification protocol.....	36
Figure 2.3.	HAV 3C protease purification and expression gel.....	44
Figure 2.4.	HSQC spectrum of specifically labelled HAV 3C using auxotrophic cell CT16	45
Figure 2.5.	Phenylalanine metabolism pathway.....	49
Figure 2.6.	HSQC spectrum of specifically labelled HAV 3C using prototrophic cell.....	50
Figure 3.1.	pET3a vector map.....	54
Figure 3.2.	pTrcHis2 vector map.....	56
Figure 3.3.	Peptide substrate of 3C protease EALFQ-pNA.....	60
Figure 3.4.	pET-HRV14 3C protease expression system.....	63
Figure 3.5.	pTrcHis2-HRV14 3C protease expression system.....	63
Figure 3.6.	HRV14 3C expression result.....	69

Figure 3.7. Adaptation of BL21(DE3) cells..... 71

Figure 3.8. HRV14 3C purification at pH 8.3..... 72

Figure 3.9. HRV14 3C purification at pH 6.8..... 73

Figure 3.10. Expression and purification from pTrc system..... 75

Figure 3.11. HRV14 3C protease colorimetric activity assay..... 76

Figure 3.12. HRV14 3C protease 1-D NMR spectrum..... 78

List of Abbreviations

Amino acids:

A (Ala)	=	L-alanine
C (Cys)	=	L-cysteine
D (Asp)	=	L-aspartic acid
E (Glu)	=	L-glutamic acid
F (Phe)	=	L-phenylalanine
G (Gly)	=	L-glycine
H (His)	=	L-histidine
I (Ile)	=	L-isoleucine
K (Lys)	=	L-lysine
L (Leu)	=	L-leucine
M (Met)	=	L-methionine
N (Asn)	=	L-asparagine
P (Pro)	=	L-proline
Q (Gln)	=	L-glutamine
R (Arg)	=	L-arginine
S (Ser)	=	L-serine
T (Thr)	=	L-threonine
V (Val)	=	L-valine
W (Trp)	=	L-tryptophan
Y (Tyr)	=	L-tyrosine

2D	=	two-dimensional
3D	=	three-dimensional
BCA	=	bicinchoninic acid
BSA	=	bovine serum albumin
CAT	=	chloramphenicol acetyltransferase
CAV	=	Coxsackie A virus
CBV	=	Coxsackie B virus
cDNA	=	complementary DNA
CM	=	carboxymethyl
DEAE	=	diethylamino ethyl
ddH_2O	=	double distilled H_2O
DNA	=	deoxyribonucleic acid

DSS	=	3-(trimethylsilyl)-1-propanesulfonic acid
DTT	=	dithiothreitol
E. coli	=	*Escherichia coli*
EDTA	=	ethylenediaminetetraacetic acid
EMCV	=	encephalomyocarditis virus
FMDV	=	foot and mouth diseases virus
HAV	=	hepatitis A virus
HRV	=	human rhinovirus
HPLC	=	high performance liquid chromatography
HSQC	=	Heteronuclear single quantum correlation spectroscopy
IPTG	=	isopropyl β-D thiogalactopyranoside
kD	=	Kilodalton
kb	=	Kilobase
L	=	Litre
LacI^q	=	*Lac* repressor gene
LB	=	Luria-Bertani medium
mg	=	Milligram
mL	=	Millilitre
mM	=	Millimolar
mm	=	Millimeter
ms	=	Millisecond
NMR	=	nuclear magnetic resonance
nm	=	Nanometer
NOE	=	nuclear overhauser effect
OD	=	optical density
PEM	=	polyethylenimine
pI	=	isoelectric point
RNA	=	ribonucleic acid

rpm	=	rotations per minute
s	=	second
SAR by NMR	=	structure activity relationship by NMR
SDS-PAGE	=	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SP	=	sulphopropyl
TOCSY	=	total correlation spectroscopy
TB	=	Terrific Broth
µL	=	Microlitre
µs	=	Microsecond
V	=	volt

Chapter 1

Introduction

Despite significant advances in medical technology and vast improvements in sanitation and water quality, viral infections continue to be a major worldwide health problem. Influenza, AIDS, hepatitis (A, B or C) are but a few examples of deadly viral scourges that afflict hundreds of millions of people. Unlike bacteria, viruses are not living creatures and so they are not susceptible to the usual drugs or treatments that would arrest bacterial growth and metabolism. Rather, viruses are obligate parasites that surreptitiously co-opt their host's metabolic infrastructure and reproduce within the protected confines of the host's cell. This concealed replication allows viruses to rapidly grow in number without being detected by the host's immune defenses. Typically viral replication continues until the infected cell bursts and releases thousands of viral particles to infect neighboring cells. This process of infecting and bursting host cells is what leads to the pathogenicity associated with most viral infections.

In many cases, the only way of stopping a viral infection is to kill the host's infected cells before the virus has a chance to reproduce inside them. Consequently many antiviral drugs, such as nucleoside analogs and DNA chelators (Brinkman et al., 2000; Pan-Zhou et al., 2000), are toxic both to the virus and to the host. Furthermore, because it is so difficult to completely

eliminate all viral particles from the host, many antiviral therapies have to be taken for extended periods of time simply to prevent a recurrence of the infection. While preventative measures for many viral infections do exist (vaccines for small pox, chicken pox, mumps and rubella for instance) prevention is not always an option. Indeed it is still a major research goal of many pharmaceutical chemists to develop an arsenal of far more effective and fundamentally less toxic anti-viral drugs that can treat viral infections after they occur.

Recently, the use of protease inhibitors as anti-viral therapies has shown considerable promise (Mills et al., 1996). Because many viruses reproduce using a specific protease or proteases to assist in the viral assembly process, this suggests that the development of protease inhibitors could allow very effective drugs to be developed that are toxic only to the virus of interest. In light of these promising developments we have turned our attention to studying the proteases found in a class of viruses, called the picornaviridae, that are known to cause a number of serious diseases in livestock and humans. This thesis is concerned with the isolation and characterization of two picornaviral proteases: the hepatitis A 3C protease and the human rhinovirus type 14 3C protease. It is hoped that this work will lead to more detailed studies of the structure and function of these proteases and eventually to the development of safe and effective protease inhibitors to treat hepatitis A and the common cold.

1.1. Picornaviruses

1.1.1. The Picornavirus Family

The picornavirus family is one of the most diverse (more than 200 serotypes) and oldest known group of viruses (Roeckert et al., 1996). Picornaviruses are the causative agent for a variety of diseases known to infect humans and livestock including hepatitis A, poliomyelitis, the common cold and foot and mouth disease (FMDV). All picornaviruses share the same common structural features including (1) a small positive sense RNA of between 7.2 kb (Human Rhinovirus type 14) and 8.5 kb (FMDV); (2) a long (600-1200 bases) 5' untranslated region and a short 3' untranslated region (50-100 bases) important for protein translation, RNA synthesis and viral virulence; (3) a genome that codes for a single "polyprotein" of 2100-2400 residues; (4) a 5' end that is modified by a small covalently attached basic protein VPg (~ 23 residues); (5) a 3' end that has a polyadenylation signal, similar to eukaryotic mRNA and (6) a protein capsid composed of 60 protomers (VP1, 2, 3 and 4) which forms a densely packed icosahedral shell (Cann, 1997).

Based on their physical properties, serological relatedness and nucleotide sequence, picornaviruses have been classified into 6 genera (Stanway et al., 2000): (1) enteroviruses (including poliovirus, Coxsackie viruses and echoviruses) (2) apthoviruses, (3) cardiovirus, (4) parechoviruses, (5) hepatitis A virus (with just a single serotype) and (6) rhinoviruses.

The poliovirus causes poliomyelitis, which results in the withering of limbs and crippling paralysis. Records of polio date to 1400 BC as determined from Temple recorders found in Egypt (Fenner et al., 1988). Prior to the development of the Salk and Sabin vaccines, polio was one of the most common causes of disability in the industrial world (Roeckert et al., 1996).

Coxsackie viruses come in two groups: Coxsackie A (CAV) and Coxsackie B (CBV). CAV infections are primarily associated with respiratory disorders, conjunctivitis and rashes. CBV infections tend to be more severe and will commonly lead to meningitis, encephalitis and myocarditis. There is no known treatment.

Foot and mouth disease virus (FMDV) is an echovirus and is the causative agent of foot and mouth diseases (FMD). It is an acute systemic disease of cloven-hooved animals, including cattle, swine, sheep and goats. FMD is one of the most feared livestock diseases due to the broad range of species affected and high rate of infectivity. A single outbreak of FMD may cost millions of dollars (Edouard et al., 1990). FMD can be controlled by vaccination (inactivated vaccine) or by the slaughter of infected animals.

The cardiovirus family is represented by a single serotype called encephalomyocarditis virus (EMCV). It can infect a wide range of mammals, from mice to humans to elephants. There are two clinical manifestations: acute myocarditis and reproductive failure. The clinical signs include lethargy,

inappetence, trembling, staggering, paralysis, poor conception rates and embryo resorption.

Parechoviruses have only recently been recognized as an isolated genus in the picornavirus family. They have been found to cause infections in the central nervous system. The main clinical symptoms of a parechovirus infection include gastroenteritis, respiratory problems, pneumonia, encephalitis and myocarditis (Stanway et al., 2000).

The focus for this particular thesis is on the last two picornaviruses: the hepatitis A virus (HAV) and the human rhinovirus (HRV).

1.1.2. Hepatitis A Virus

As its name suggests, the hepatitis A virus or HAV attacks the liver and is the most common cause of hepatitis worldwide. Hepatitis A is a relatively mild illness characterized by a sudden onset of fever, malaise, nausea, and abdominal discomfort, followed by a bout of jaundice. However, hepatitis A is also a significant risk factor for liver cirrhosis and hepatocellular carcinoma (Stanway et al., 2000). In U.S., type A hepatitis accounts for 38% of all reported hepatitis cases or about 22,700 cases per annum. The hepatitis A virus is primarily transmitted through fecally contaminated food and water. Consequently hepatitis A is particularly predominant in countries with poor water sanitation and purification systems (Lycke et al., 1983). Despite its widespread prevalence, most adults develop a life-long immunity against reinfection. Recently, a number of

vaccines for HAV have become available (Andre et al., 1992; De Febres et al., 1999).

1.1.3. Human Rhinovirus

Human rhinovirus (HRV) is the causative agent for the common cold (acute afebrile upper respiratory disease). The common cold is perhaps the most frequently contracted and geographically widespread disease in the world. Although not always serious, this disease causes the loss of more than 200 million work and school days each year in United States alone (Renato et al., 1988). The antigenic features on the surfaces of HRV are complex, and so far more than 100 serotypes have been identified (Kapikian et al., 1967; Hamparian et al., 1987). The high level of infectivity for the common cold is partially due to this high antigenic variability. Interestingly, the infection of a single rhinoviral serotype can lead to the immunity to this serotype for ~ 2 years, but will not confer immunity to other serotypes (Renato et al., 1988).

1.2. Picornaviral 3C Proteases

1.2.1. Viral Protease and Polyprotein Maturation

When a picornavirus infects a cell it sheds its protein coat and releases its RNA genome which is translated by the host cell's machinery into a polyprotein of ~ 250 kD. This polyprotein contains all the proteins needed for processing and

maturation. A map of the picornaviral polyprotein is shown in Figure 1.1. The P1 region contains the precursor viral capsid proteins, while the P2 and P3 regions contain the precursors of functional proteins important for viral RNA replication. The viral maturation process can be divided into 3 steps, which are mainly carried out by the viral proteases 2A and 3C. The primary cleavage is carried out cotranslationally by the 2A protease which separates the P1 region from the P2-P3 region by cleaving its own amino terminus. Subsequently, the majority of the cleavages within P1, P2 and P3 are carried out by the 3C protease. The final processing step in viral maturation is to cleave VP4 from VP2 (viral capsid proteins), whose mechanism is not very clear.

1.2.2. Picornaviral 3C Protease Structure

Picornaviral 3C proteases are small, 20 kD proteins that typically cleave at Gln-Gly recognition sites. The 3C proteases are not only responsible for viral maturation and polyprotein processing, but they are also involved in viral replication as an RNA-binding protein (Porter, 1993). 3C proteases also appear to regulate host cell function by cleaving important cellular proteins during infection (Clark et al., 1991; Yalamanchili et al., 1997). To date, X-ray crystal structures from three genera of picornaviral 3C protease have been solved: hepatitis A (Allaire et al., 1994), human rhinovirus type 14 (Matthews et al., 1994), and poliovirus (Mosimann et al., 1997). The human rhinovirus type 14 3C protease crystal structure (Matthews et al., 1994) reveals the presence of two topologically

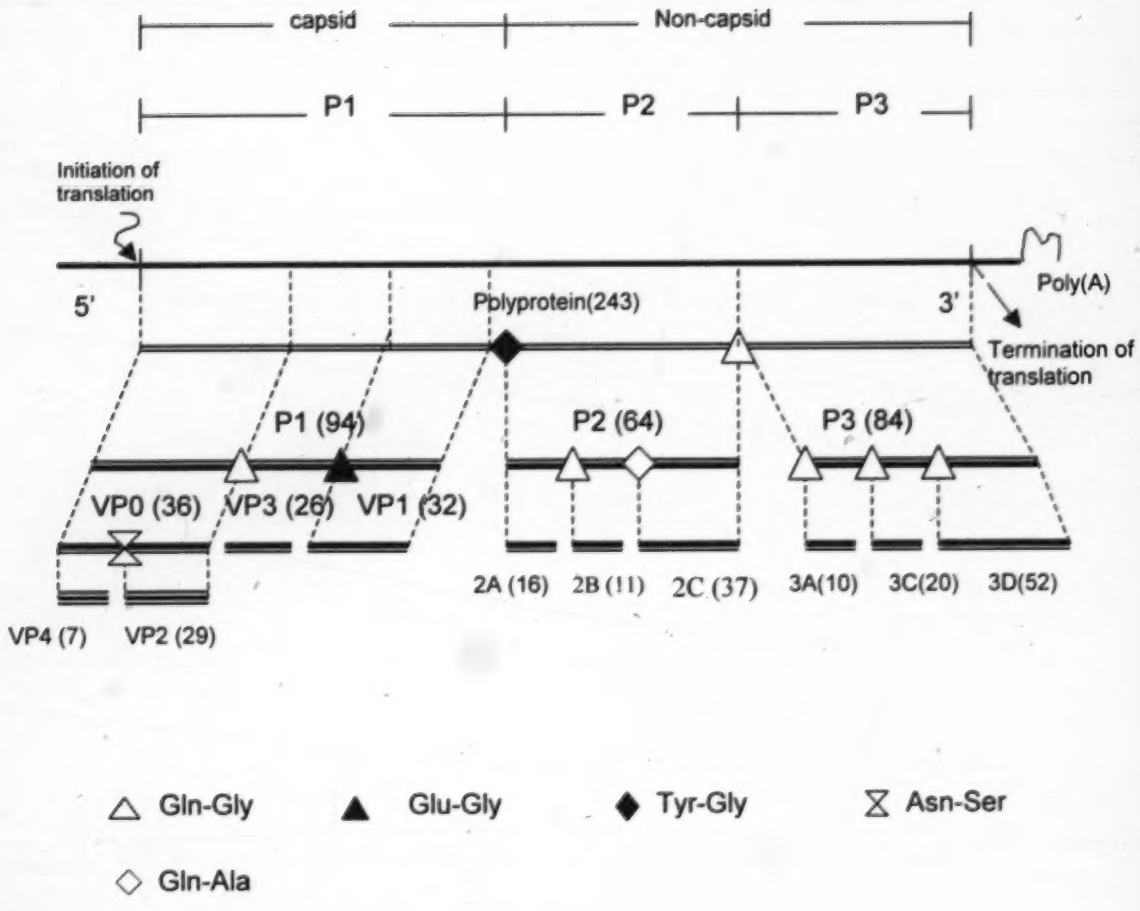


Figure 1.1. Proposed HRV-14 protein cleavage scheme. Numbers in parentheses indicate the molecular masses of polypeptides in kilodaltons (Cordingley et al., 1990).

equivalent, six-stranded β-barrels. These two domains combine to create an extended shallow groove for substrate binding. The catalytically important residues generally form a linked cluster of amino acids.

The 3C proteases of both HRV and HAV have been identified as cysteine protease because mutation of the proposed catalytic cysteine leads to an inactive protease *in vitro* (Kay et al., 1990; Cordingley et al., 1989; Gosert et al., 1997). Despite the use of a cysteine as a catalytic nucleophile, the 3C proteases are more similar to trypsin-like serine proteases in structure, than papain-like cysteine proteases (Matthews et al., 1994; Malcolm et al., 1995).

1) HRV 14 3C protease. The catalytic cluster in HRV14 3C protease is composed of Cys146, His40 and Glu71 (Figure 1.2). The highly conserved sequence Gly-X-Cys-Gly-Gly serves to position the active site cysteine for nucleophilic attack on the substrate's carbonyl carbon of the scissile bond and to orient the backbone amino groups of the Gly144 and Cys146 to form an "oxyanion hole" for stabilizing of the proposed tetrahedral transition state (Matthews et al., 1994). Asn164 in the HRV 3C protease has also been found to be critical for substrate recognition and binding (Cox et al., 1999). When this residue was deamidated to an aspartic acid, the protein's isoelectric point changed from 9.0 to 8.3 along with a concomitant 10-fold reduction in activity. Deamidated 3C protease also exists *in vivo*, and its level of abundance seems to play a role in the regulation of 3C protease activity (Cox et al., 1999).

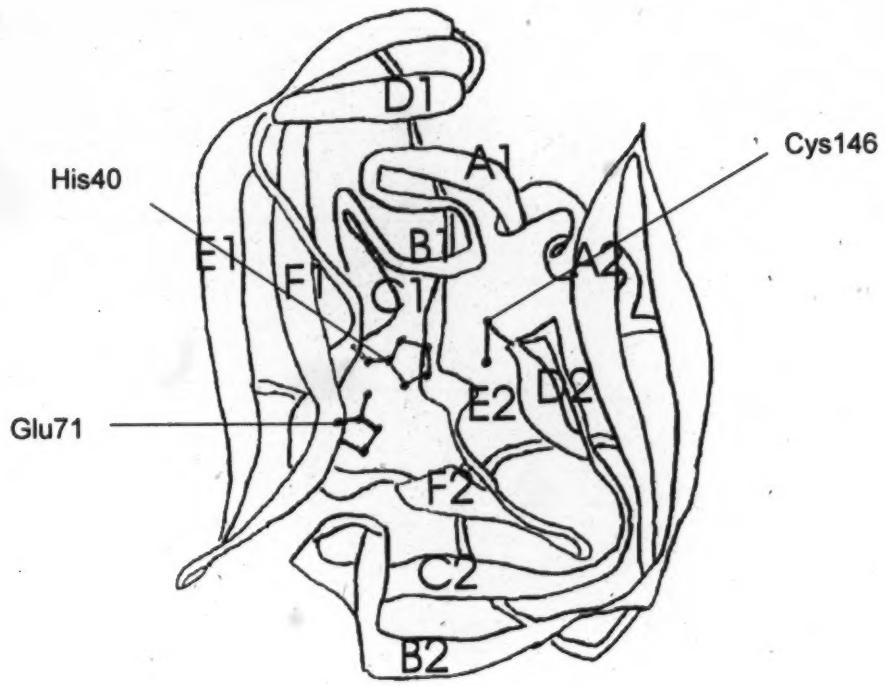


Figure 1.2. Overall architecture of HRV14 3C protease. The side chain positions shown are three residues of active site, Cys146, His40 and Glu71. (Matthews et al., 1994)

For HRV 14 3C cleavage activity, there are three different native cleavage sites in polyprotein: Gln/Gly, Glu/Gly and Gln/Ala. But for the synthetic peptide substrates they require an absolute Gln/Gly scissile bond. The most efficient cleavage site consists of Gln-Gly-Pro (Cordingley et al., 1989; Orr et al., 1989). The studies made by Cordingley (1989) determined a minimal substrate of 6 amino acids (TLFQ/GP) for the 3C protease cleavage reaction. But maintenance of P5 position (ETLFQ/GP) will increase the cleavage efficiency by 10 fold. Positions P1, P4, P1' and P2' are critical to the activity. Long (1989) defined the minimum and highest efficient substrate as P5-Val/Thr-P3-P2-Gln-Gly-Pro.

High-resolution X-ray crystal structures for 10 serotypes of human rhinovirus 3C proteases bound to peptide-based aldehyde inhibitors have revealed that the substrate binding sites are highly conserved among different viral serotypes (Matthews et al., 1999). Since the HRV 3C proteases have similar binding pockets, a single anti-viral drug could potentially be used to inhibit all serotypes.

2) HAV 3C protease. Just like HRV 3C protease, HAV 3C protease consists of two antiparallel β-barrel domains as well, with the active site and substrate binding region lying in the surface groove between the two β-barrel domains (Bergmann et al., 1997). In HAV 3C protease, the catalytic cluster is composed of Cys172 as nucleophile and His44 as general base (Figure 1.3). His191 primarily determines the 3C cleavage specificity for glutamine residues. Asp84 does not play an important role in the catalysis because its side-chain is turned 180° away

from the His44 side-chain (Seipelt et al., 1999). Tyr143 appears to maintain the orientation of the imidazole ring of His44 (Bergmann et al., 1997). The catalytic dyad is formed by Cys172 and His44, while the oxyanion hole is composed of Pro169, Gly170, Met171 and Cys172, which is above the dyad and to the right (Bergmann et al., 1997). HAV 3C protease has four pockets (S1 - S4) to bind four substrate positions (P1 - P4). The S1 pocket consists of backbone of Ala 193, Gly194, Gly167 and Leu168 and side chain of Leu199, Met171 and His191. The S2 pocket is formed by Tyr143, His145 and Ala193. The S3 pocket is composed of Val144 and Lys146. The S4 pocket is made of the side chains of Ala141, Val144, Ala198 and Val200.

For the cleavage site in HAV 3C substrate, glutamine or glutamic acid is required at P1 position, but a variety of residues are accepted at P1' position, including methonine, valine, alanine, serine, glycine, and arginine (Seipelt et al., 1999). Possibly this is because that a bulge in β-strand bI allows several carbonyl groups to act as hydrogen bond acceptors for P' substrate residues. Gln in P1 position and Leu in P4 position are determined to be important for the cleavage specificity (Jewell et al., 1992).

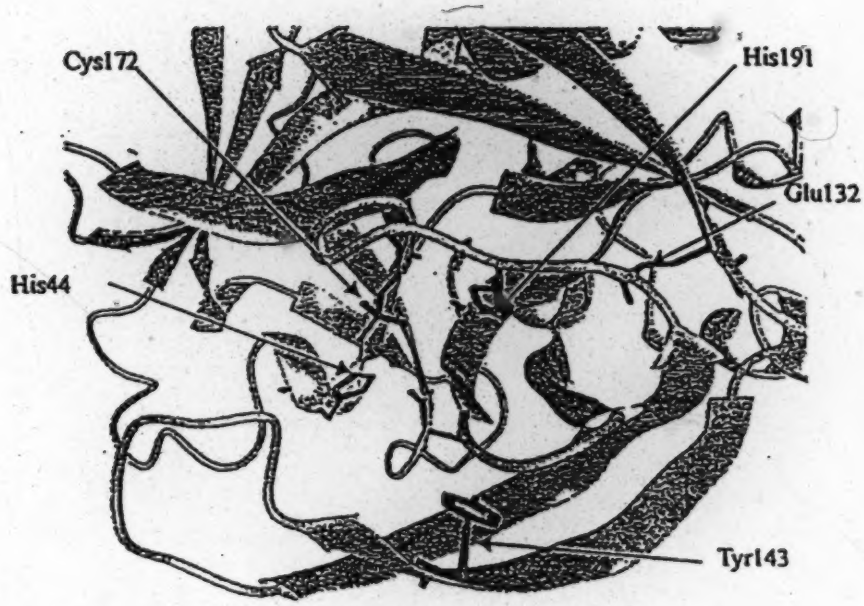


Figure 1.3. Active site residues of HAV 3C protease.
(Reproduced by permission of M.S. Watson).

1.3. Structure-Assisted Design of 3C Protease Inhibitors by X-ray Crystallography

Over the past six years, a number of potentially effective 3C protease inhibitors have been designed with the assistance of X-ray crystallography. Some of these compounds have already entered clinical trials including a number of reversible inhibitors, such as tripeptide-derived molecules incorporating C-terminal ketone electrophiles (Dragovich et al., 2000). However, irreversible inhibitors appear to offer more promise because of the unusual properties of the 3C protease binding site. In particular, the specificity pockets for substrate (or inhibitor) recognition in 3C proteases tend to be quite shallow, and binding determinants are widely dispersed over large surface areas. Consequently it is difficult for a small inhibitor molecule to bind structurally well to inhibit the interaction of 3C protease with its authentic viral polyprotein substrate. However, it is known that inhibitor potency can be improved by forming a covalent adduct with the reactive cysteine at the active site of the protease (Matthews et al., 1999). This approach involves chemically modifying a specific noncovalent substrate analog with an electrophilic anchoring group that reacts covalently with the active site nucleophile. The specific substrate analog is usually a tri- or tetrapeptide (Wang, 1999). These peptides increase the affinity of the inhibitor for the

protease, while the electrophilic component moiety interacts with the active site nucleophile and inhibits the protease cleavage activity.

For HRV 3C protease inhibitors, the electrophile could be an aldehyde, a Michael acceptor or a bromomethylketone. One compound (AG7088, Figure 1.4) developed at Agouron Pharmaceutical Inc. (San Diego, CA, USA) actually displayed inhibitory activity against all 48 rhinovirus serotypes tested so far, and is currently in clinical trials (Patick et al., 1999).

For inhibitors of HAV 3C protease, a combination of a fluoromethylketone (FMK) and peptide that mimics the natural peptide specificity such as Acetyl-Leu-Ala-Ala-Gln-FMK (Figure 1.5) has been shown to be an effective inhibitor to the HAV 3C protease both *in vitro* and *in vivo* (Morris et al., 1997). Other potential inhibitors under studies include α, β –unsaturated carboxylesters, β- and γ-lactones, lactams, isatins (2,3-dioxindoles), and triterpene sulfates (Skiles et al., 1990; Brill et al., 1997).

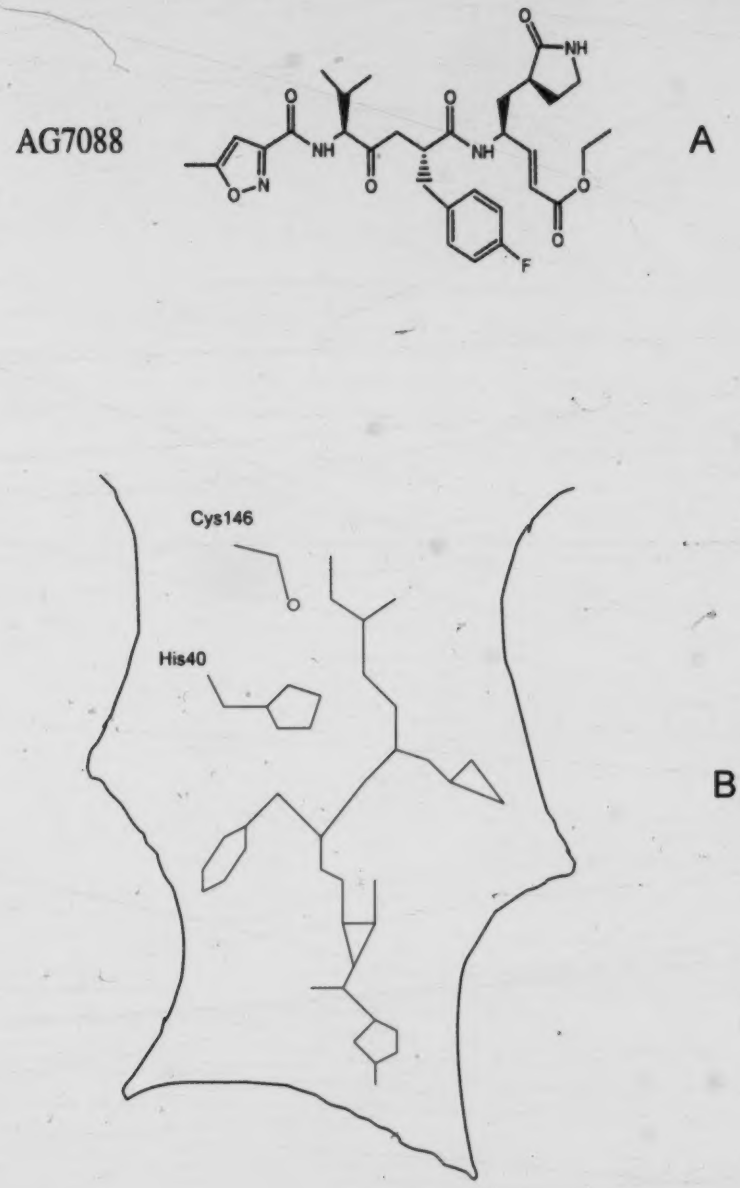


Figure 1.4. A. HRV 3C protease inhibitor AG7088 (Matthews et al., 1999).
B. AG7088 interacts with the active site of 3C protease.

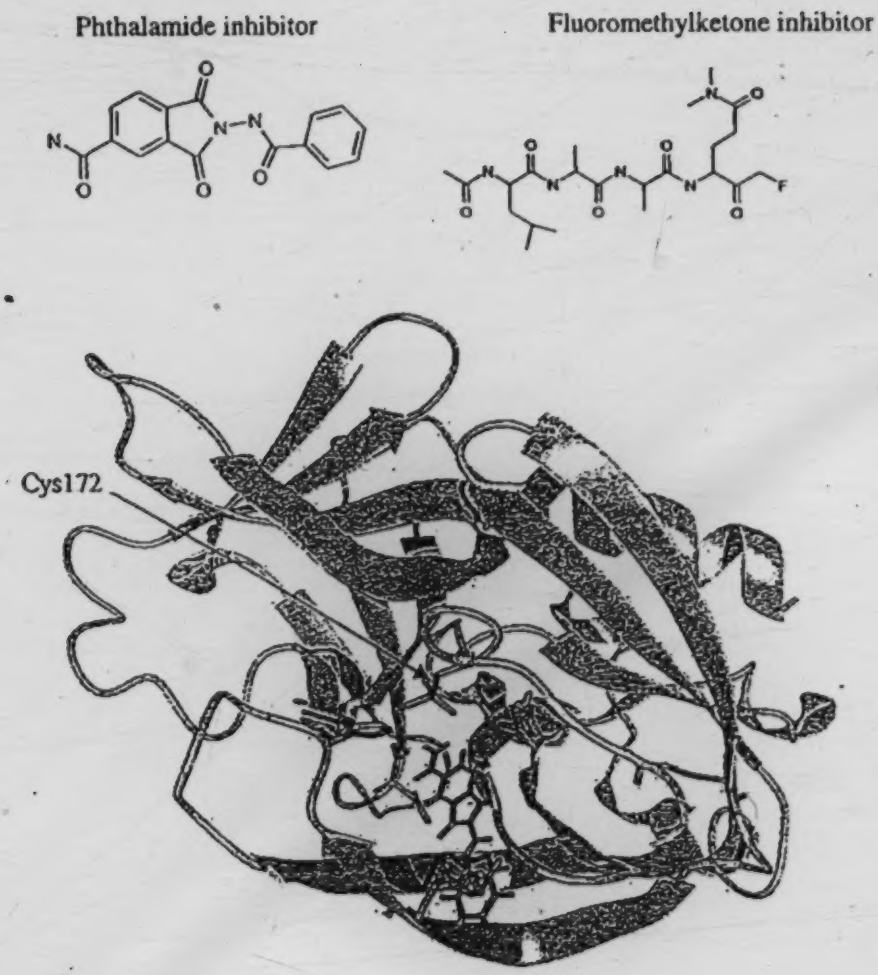


Figure 1.5. Examples of inhibitors to HAV 3C protease. The pthalamide inhibitor is positioned in HAV 3C protease's active site such that its aldehyde reacts with the nucleophilic thiol of Cys172. (Reproduced by permission of M. S. Watson).

1.4. Structure-Assisted Design of 3C Protease Inhibitors by NMR Spectroscopy

1.4.1. Screening Protein Inhibitors by NMR

Traditional methods of drug discovery are based on the manual screening of chemical libraries or natural products to identify lead compounds. This process is normally followed by further optimization of inhibitory activity using "intuitively" derived chemical modifications of the lead compound. Without the knowledge of the target protein structure or the nature of the binding site, this kind of screening and optimization can be time consuming and inefficient. Structure-assisted design, a rational design approach based on using structural data to help screen and optimize lead compounds is one approach that can often improve the speed and efficiency of the drug discovery process (Bowen et al., 1993; Wlodawer et al., 1998). Structure assisted drug design makes use of such tools as X-ray crystallography, NMR spectroscopy and computer analysis to make the process substantially more efficient and more automated than traditional approaches. There are at least two different NMR techniques that can be applied to discover high affinity protein ligands: Affinity NMR (Chen et al., 1999) and SAR-by-NMR (Shuker et al., 1996).

Affinity NMR, also called diffusion-edited NMR, is a high throughput technique that identifies compounds which bind directly to a large receptor

molecules. As such, it offers great promise in interpreting inhibitor-screening results. Affinity NMR distinguishes active ligands, which effectively bind to a protein, from non-active ones by detecting the change in the ligand diffusion constant. The diffusion rate of a small ligand changes significantly when it binds to a large molecule. The difference in diffusion coefficients between bound and unbound ligands can be routinely detected by NMR, thereby allowing one to rapidly detect those ligands that bind to the target molecule (Chen et al., 1999).

SAR-by-NMR (structure-activity relationships by NMR) is another high throughput NMR technique that can be used to discover high affinity ligands for proteins (Shuker et al., 1996; Hajduk et al., 1997). It utilizes two-dimensional ^{15}N-heteronuclear single-quantum correlation (^{1}H-^{15}N-HSQC) spectra of uniformly labelled proteins to identify lead compounds. The ligand binding location and ligand affinity can be determined by observing ^{15}N or ^{1}H amide chemical shift changes in the HSQC spectra. More specifically SAR-by-NMR involves five straightforward steps: (1) a library of low molecular weight compounds is screened by analyzing chemical shift differences in HSQC spectra to identify molecules that are able to bind to the protein; (2) once a candidate molecule has been identified, analogs of the candidate are screened to optimize the binding to this site; (3) another ligand is sought to bind to a site nearby to the first one; (4) the binding site of the second ligand is approximately located based on the chemical shift change on the NMR spectra, the second ligand is also optimized and (5) when the two "lead" fragments have been selected, their

location and orientation are determined by standard NMR procedure. Based on this information, a compound linking the two "lead" ligands would be designed and synthesized. The success of SAR-by-NMR strongly depends on the fact that ^{1}H-^{15}N HSQC spectra can be rapidly collected (15 minutes) making this process ideal for screening and characterizing large numbers of compounds. Because of the importance of ^{15}N labelling in protein-ligand analysis and particularly in the 3C protease work described in this thesis, it is perhaps appropriate to discuss this subject in more detail.

1.4.2. ^{15}N Nuclear Magnetic Resonance

All proteins contain nitrogen, and because nitrogen is known to exhibit considerably more chemical shift dispersion than either carbon or hydrogen, it has long been an objective of NMR spectroscopist to employ ^{15}N NMR in detailed studies of peptides and proteins (Witanowsksi et al., 1985; Levy et al., 1979). Prior to 1985, however, the application of ^{15}N NMR to biological systems was somewhat limited by its low natural abundance (0.36%), and the inherent insensitivity of the ^{15}N nucleus ($\gamma^{15}N/\gamma^{1}H$= - 0.101). However, with the near simultaneous introduction of powerful inverse detection methods (INEPT, HMQC) (Morris et al., 1979; Henry et al., 1990; Kay et al., 1989), higher magnetic field strengths, and improved methods for isotopic labelling (LeMaster et al., 1985), the whole field of ^{15}N NMR changed almost overnight. Indeed, ^{15}N

NMR is now one of the most powerful and widely used spectroscopic tools for studying the structure and dynamics of biomolecules (Dekker et al., 1991; Yamazaki et al., 1992). An enormous range of multinuclear and multidimensional NMR experiments are now available to exploit the favorable spectroscopic qualities of ^{15}N, including ^{1}H-^{15}N HSQC, ^{1}H-^{15}N TOCSY-HSQC, ^{1}H-^{15}N NOESY-HSQC, HNHA, HNCACB and HNCO experiments (Sun et al., 1997; Salzmann et al., 1998; Moy et al., 1995). These experiments are now part of the standard repertoire that biomolecular NMR spectroscopists use to assign and characterize peptides, proteins and polynucleotides.

While advances in NMR methodology and instrumentation have played a key role in the development of ^{15}N NMR, the importance of molecular biology to this field cannot be ignored. It is only through advances in molecular biology that it is now possible to prepare the tens or hundreds of milligrams of isotopically labelled protein necessary to conduct the heteronuclear NMR experiments described above. In the case of ^{15}N labelling, it is often necessary to use bacterial expression systems that have been specially modified to carry the gene of interest through recombinant DNA techniques. These recombinant bacterial strains must also be sufficiently robust to grow on defined minimal media supplemented with a ^{15}N source.

Over the past decade two types of ^{15}N labelling techniques have evolved: uniformly labelling and residue-specific labelling. Both techniques have very specific media and cell strain requirements and both have their advantages and

disadvantages. Because of its importance for this thesis we shall focus mostly on describing residue-specific labelling.

1.4.3. Selective ^{15}N Labelling of Proteins by Residue Type

1) General Background. Selective isotopic labelling is an ^{15}N enrichment strategy in which one or two types of ^{15}N amino acids are selectively incorporated into the protein of interest by their judicious substitution into the growth media. Selective isotopic labelling offers a means of simplifying the inherent complexity of NMR spectra by reducing the number of visible peaks from hundreds to only a dozen (Otting et al., 1986; Bax et al., 1987; Griffey et al., 1987). This simplification can be an invaluable aid in assigning residues by amino acid type, in monitoring key amino acids or in rapidly assessing spectral changes upon solvent or solute perturbations (Opella et al., 1987; Oas et al., 1987). Since the 1970's dozens of proteins have been selectively labelled to facilitate NMR studies, including hemoglobin (Lapidot et al., 1976), α-lytic protease (Bachovchin, 1985), T4 lysozyme (Muchmore et al., 1989), *E. coli* thioredoxin (LeMaster et al., 1985) and λ *cro* repressor (Leighton et al., 1987).

It is now clear that there are at least two ways of performing selective isotopic labelling: (1) the use of a prototrophic cell line, with an excess supply of the desired ^{15}N amino acids in the growth medium or (2) the use of a bacterial strain with special genetic deficiencies, i.e., auxotrophic cell line.

2) Selective ^{15}N Labelling Using Prototrophic Cell Strains. When an excess amount of a labelled amino acid is added to the growth media, the endogenous synthesis of this type of unlabelled amino acids is inhibited through kinetic suppression of the corresponding amino acid synthetases and transferases. This simple approach (Senn et al., 1987) is not particularly reliable because isotopic dilution or isotopic "scrambling" (i.e. incorporation of the ^{15}N label into undesired sites) may occur. Isotopic dilution is a problem manifested by a diminishment in the isotopic abundance of the chosen isotopic label. Normally one strives for near 100% incorporation of ^{15}N labelled amino acids into proteins, however in practice this often proves to be difficult. Generally, the presence of endogenous transaminases (enzymes which swap amino groups in amino acids) will lead to the removal of ^{15}N amino groups from labelled amino acids and their concomitant replacement with ^{14}N amino groups. This amino group swapping will lead to measurable ^{15}N isotopic dilution. Isotopic dilution can also occur in prototrophic bacteria where endogenous synthesis of unlabelled amino acids leads to a reduction in the proportion of added ^{15}N labelled amino acids. The use of auxotrophic bacterial strains is the best route to solve this problem (Muchmore et al., 1989).

Unlike isotopic dilution, isotopic scrambling is mainly caused by the metabolic conversion of the intended ^{15}N labelled amino acids to other ^{15}N amino acids through general nitrogen pool or through aminotransferase catalyzed nitrogen exchange. Aminotransferases catalyze the reversible transfer of amino

groups from amino acids to the α-keto precursors of other amino acids. *E. coli* is known to have four types of aminotransferases and their presence can lead to adverse effects on the quality and specificity of selective labelling (McIntosh et al., 1990). To solve the above two difficulties, a number of auxotrophic cell strains have been developed.

3) Selective ^{15}N Labelling Using Auxotrophic Bacterial Strains. To selectively label amino acids that are either the precursors of other amino acids or the substrates of aminotransferases, auxotrophic cells must be used to achieve high levels of specific labelling. Based on the type of metabolic gene defect, auxotrophic strains can be classified into two categories: 1) cells that are auxotrophic for certain amino acids due to amino acid synthetase deficiencies, such as the genotypes *glnA*-, *argH*- and *lysA*- and 2) cells that are auxotrophic due to an amino transferase deficiency, such as the genotypes *aspC*-, *tyrB*-, *ilvE*- and *avtA*-. The first category applies to isotopic labelling with Arg, Cys, Gln, Gly, His, Ile, Lys, Met, Pro and Thr (Waugh, 1996). The second category applies to isotopic labelling with Ala, Asp, Glu, Leu, Phe, Trp, Tyr, Ser and Val (Waugh, 1996). If an amino acid is the substrate of more than one aminotransferase, a cell strain should have deficiencies in all of the associated amino transferase genes to ensure full auxotrophy.

To construct a genetic lesion in a cell strain, defective genes must be transformed from the chromosome in one strain to another. Two methods are

currently available, generalized transduction and conjugational transfer (Miller, 1972). In both methods the motile defective gene is inserted into the recipient cell chromosome by homologous recombination. In generalized transduction, bacteriophage P1 is used to transfer the desired gene. Conjugational transfer relies on the mating of donor and recipient cells via the F-pilus, whereupon part of the chromosome is transferred. The selection of transformed cells depends upon the linked selectable transposon marker, which is usually an antibiotic resistance gene. The auxotrophic cell described in this thesis was constructed by generalized transduction (Waugh, 1996).

Chapter 2

Residue Specific ^{15}N labelling of HAV 3C protease

2.1. Introduction

The Hepatitis A virus 3C (HAV 3C) protease is critical to the Hepatitis A virus life cycle. Developing inhibitors that could specifically arrest the activity of this cysteine protease could lead to significant improvements in the treatment of Hepatitis A. One approach for accelerating the development of such protease inhibitors is to use structure-based drug design principles, where the 3D structure of the protease is used to design, refine and test lead compounds. In this regard, structural studies of the HAV 3C protease have been ongoing for more than 8 years (Malcolm et al., 1992; Allaire et al., 1994; Bergmann et al., 1997 & 1999). The 1.8 Å X- ray structure of the HAV 3C protease (PDB ID 1QA7) reveals that the protein has a two-domain β-barrel structure, similar to that seen in chymotrypsin-like serine proteases. To complement these X-ray studies, our laboratory has been working on the NMR assignments and solution structure of HAV 3C protease. It is hoped that this NMR approach will allow a wide variety of inhibitors to be rapidly screened and the bound structures determined using techniques like SAR-by-NMR (Shuker et al., 1996) or NOE-pumping (Chen et al., 1999).

Key to using NMR in structure-based drug design is the need for a full set of backbone and side chain assignments. Efforts have been ongoing to complete these assignments using a variety of heteronuclear approaches (HNCACB, HNCO,

HNCOCA, HSQC, HSQC-TOCSY) applied to uniformLy $^{13}C/^{15}N$ labelled material (Watson, 2000). To complement these approaches we have also adopted site specific labelling to assist with residue-type identification. Using a prototrophic strain of *E. coli* (MM294), we found that site-specific ^{15}N labels on HAV 3C protease could be generated for alanine, lysine, isoleucine, valine, leucine, and to a lesser extent, glycine (Watson, 2000). However, attempts to generate ^{15}N phenylalanine-labelled material have consistently failed.

Because many phenylalanine residues are located in key regions in this enzyme and because their identification would assist in the completion of the HAV 3C backbone assignments, we decided to investigate the use of a specially developed auxotrophic strain of *E. coli* (CT16) developed by Dr. David Waugh (Waugh, 1996). We attempted to explore whether we could successfully label and identify the phenylalanine residues in HAV 3C protease using this mutant bacterial strain. If successful, this approach could open the door to site-specific labelling attempts for other amino acids in HAV 3C and for other picornaviral proteases being studied in our laboratory.

2.2. Materials and Methods

2.2.1. Reagents and Instruments

^{15}N-Phenylalanine was purchased from Cambridge Isotope Laboratories Inc., (Andover, MA, USA). Unlabelled amino acids were obtained from Sigma-Aldrich Canada Ltd. (Oakville, ON). CM-Sepharose Fast Flow chromatography medium was purchased from Amersham Pharmacia Biotech, Pharmacia Biotech Inc.(Uppsala, Sweden). Reagents used for transformation, expression and purification are listed in Tables 2.1 to 2.3.

2.2.2. Cell Strains and Vectors

The CT16 *E. coli* cell strain was a gift from Dr. D. Waugh (Roche Research Center, Hoffman-La Roche Inc., Nutley, NJ, USA). The CT16 strain is defective in four aminotransferase genes (*ilvE*12, *tyrB*507, *aspC*13 and *avtA*::Tn5) and is auxotrophic for phenylalanine. Phenylalanine is the substrate for two aminotransferases (A and D) which are encoded by the *aspC* and *tyrB* genes, respectively.

The vector containing the HAV 3C protease coding sequence pHAV-3CEX (Figure 2.1) was a gift from Dr. B. A. Malcolm (Schering Plough Inc., NJ, USA). It is an *E. coli* compatible multicopy expression vector that produces the HAV 3C protease as a soluble and active enzyme. The expression level was found to

Table 2.1. ^{15}N-phenylalanine supplemented amino acid enriched M9 medium.

Components	Volume
1% ^{15}N-Phe	10 mL
1% Trp	10 mL
1% Tyr	10 mL
1% Asp,Glu	10 mL
1% Cys,Asn,Gln	10 mL
1% Leu,Ile,Val	10 mL
1% Met	10 mL
1% Ala,Thr,Pro,Ser	10 mL
1% Gly	10 mL
1% His	10 mL
1% Arg,Lys	10 mL
20% Glucose	20 mL
1 mM $FeSO_4$	2 mL
0.01M $ZnSO_4$	5 mL
1 M $MgSO_4$	2 mL
1 M $CaCl_2$	0.1 mL
Vitamin mixture	
d-Biotin	1 mg
Choline Chloride	1 mg
Folic Acid	1 mg
Niacinamide	1 mg
d-pantothenic acid	1 mg
Pyridoxal Hydrochloride	1 mg
Riboflavin	0.1 mg
ddH_2O	10 mL
***M9 salts**	
Na_2HPO_4	6.78 g
KH_2PO_4	3.0 g
NaCl	0.5 g
ddH_2O	200 mL
*ddH_2O	650 mL
Ampicillin	100 mg

* These solutions were autoclaved separately. The other reagents were filtered using a 0.22 μm sterile filter. All of the above components were then mixed to prepare the final growth medium.

Table 2.2. Terrific Broth Recipe

Components	Amount
ddH_2O *	896 mL
Tryptone *	12 g
Yeast Extract *	24 g
Glycerol *	4 mL
TB salts	
KH_2PO_4 †	2.31 g
K_2HPO_4 †	12.54 g
ddH_2O †	100 mL

* Mixed together, autoclaved and cooled to 60°C, then add TB salts.
† Sterilized using a 0.22 μm sterile filter.

Table 2.3. SOC transformation medium recipe

Components	Amount
Bactotryptone	2.0%
Yeast Extract	0.5%
NaCl	10 mM
KCl	2.5 mM
$MgCl_2$	10 mM
$MgSO_4$	10 mM
Glucose	20 mM

* SOB is the same as SOC except that it lacks glucose.

represent approximately 10% of total cell protein. The coding sequence for HAV 3C protease was obtained from a HAV cDNA clone isolated previously (Malcolm et al., 1992). An initiation codon and a translation enhancing "minigene" were added upstream of the 3C protease gene sequence. To improve expression levels, the combined sequence was inserted into a ptac5SOD plasmid to create the pHAV-3CEX plasmid. pHAV-3CEX is under the control of the *tac* promoter and is induced by the addition of isopropylthio-β-D-galactoside (IPTG) (See Figure 2.1).

2.2.3. Plasmid Isolation

Isolation of the pHAV-3CEX plasmid from the host MM294 cells was carried out using a QIAGEN QIAprep Miniprep Kit (QIAGEN Inc., Valencia, CA, USA). Glycerol stocks of MM294 cells harboring the pHAV-3CEX plasmid were streaked onto a LB-agar plate and incubated at 37°C for 18 hours. Sixteen colonies were inoculated individually into 3 mL of LB media each. These cultures were grown at 37°C overnight (~16 hours) with shaking at 300 rpm. A 1.5-mL aliquot was taken from each tube, transferred to microfuge tubes and centrifuged at 13,000 rpm for 1 minute. The cell pellet was resuspended in 250 µL of Buffer P1 (Resuspension Buffer, 50 mM Tris•HCl, pH 8.0; 10 mM EDTA; 100 µg/ mL RNase A). 250 µL of Miniprep Buffer P2 was added (Lysis Buffer, 200 mM NaOH, 1% SDS), and the solution gently mixed by inverting the tubes 4 to 6 times. Buffer N3 (potassium acetate, glacial acetic acid solution) was then added (350 µL)

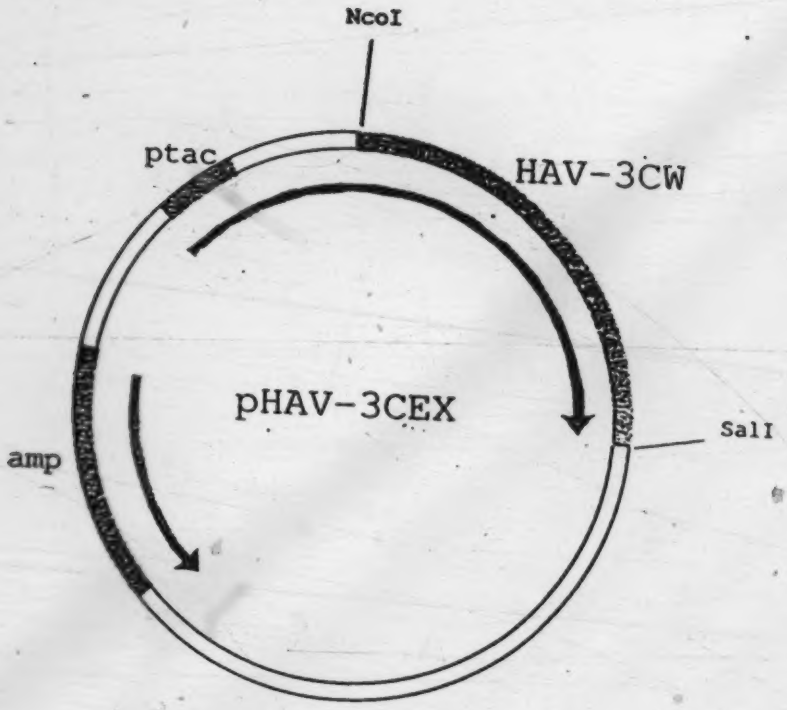


Figure 2.1. The pHAV-3CEX vector map. The NcoI-SalI fragment containing HAV3C gene was inserted into a pTac5SOD vector for expression. The HAV 3C gene is under the control of a *tac* promoter and is induced by the addition of IPTG. The selective marker is β-lactamase (ampicillin resistance).

and mixed as before. The samples were centrifuged at 13000 rpm for 10 minutes. The supernatants were loaded onto the QIAprep column by decanting the supernatant into the Qiagen spincolumn, which was then centrifuged for 60 seconds. The QIAprep columns were washed with 0.5 mL of Miniprep Buffer PB (phosphate buffer) and then 0.75 mL of Miniprep Buffer PE, centrifuged at 13000 rpm for 60 seconds after each wash. The columns were centrifuged a second time to remove excess wash buffer and then placed into a clean 1.5-mL microfuge tube. To elute the plasmid DNA, 50 μL of Buffer EB (10mM Tris•HCl, pH 8.5) was added to each QIAprep column. After standing for 1 minute at room temperature, the tubes were centrifuged for 1 minute to collect the plasmid DNA. The concentration of DNA was determined to be approximately 0.2 μg/μL using UV absorbance (OD_{260} nm).

2.2.4. Cell Transformation

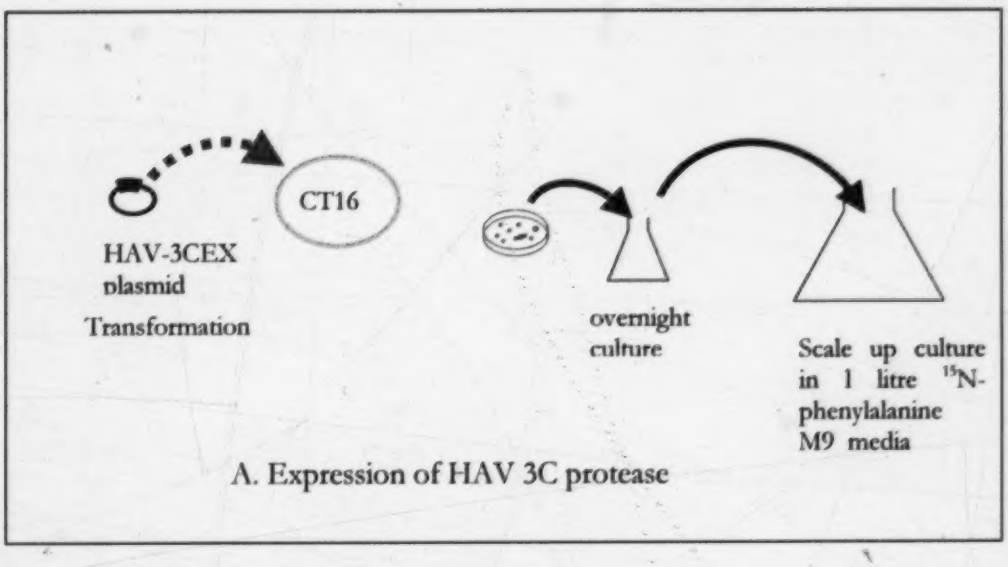
1) Preparation of Electrocompetent Cells. Glycerol stocks of CT16 cells were streaked onto a LB-agar plate and incubated overnight at 37°C. A single colony was chosen and used to inoculate 50 mL of SOB medium (See Table 2.3). This was grown at 37°C overnight in an incubator-shaker (300 rpm). Approximately 5 mL of the overnight culture was used to inoculate 500 mL of SOB medium, which was then shaken at 250 rpm (37°C) until the cell density reached an $OD_{550} = 0.8$ (~ 2.5 hours). The culture was chilled on ice before being transferred to chilled centrifuge bottles which were centrifuged for 15 minutes at 2500 g (4°C) to collect the cells.

The supernatant was decanted and the cell pellet washed twice with ice-cold sterile distilled water followed by two washes with ice-cold sterile 10% glycerol. The cells were then centrifuged at 2500 g for 15 minutes. A small amount (~ 0.5 mL) of ice-cold 10% glycerol was used to resuspend the pellet. The cells were divided into 130 μL aliquots, frozen in dry ice and stored at – 80°C.

2) Electroporation. Sterile microcentrifuge tubes, electroporation cuvettes (1mm width), electrocompetent cell stock and the DNA solution were all kept on ice. A small amount (3 μL) of the DNA solution and 40 μL of electrocompetent cells were added to a 1.5 mL microcentrifuge tube, mixed and loaded into the electroporator cuvette. This cuvette was then placed into the electroporator (model EC-100, E-C Apparatus Corp., USA) and the cells given a 1800 volt pulse for 5 ms. The cells were then rinsed out from the electrode gap using 1 mL of SOC (See Table 2.3) medium and incubated at 37°C in a shaker (250 rpm) for 1 hour. This culture was streaked onto a LB/Amp-agar plate and incubated at 37°C overnight.

2.2.5. Expression of Selectively ^{15}N Phenylalanine Labelled HAV 3C Protease

A simple schematic describing the expression and purification of HAV 3C protease is shown in Figure 2.2. A single colony was chosen from the transforming plate (see above) and used to inoculate 30 mL of LB media (containing 50 μg/mL of ampicillin). This culture was incubated at 37°C overnight with shaking at 300 rpm. 1 litre of amino acid enriched minimal media with ^{15}N-labelled phenylalanine



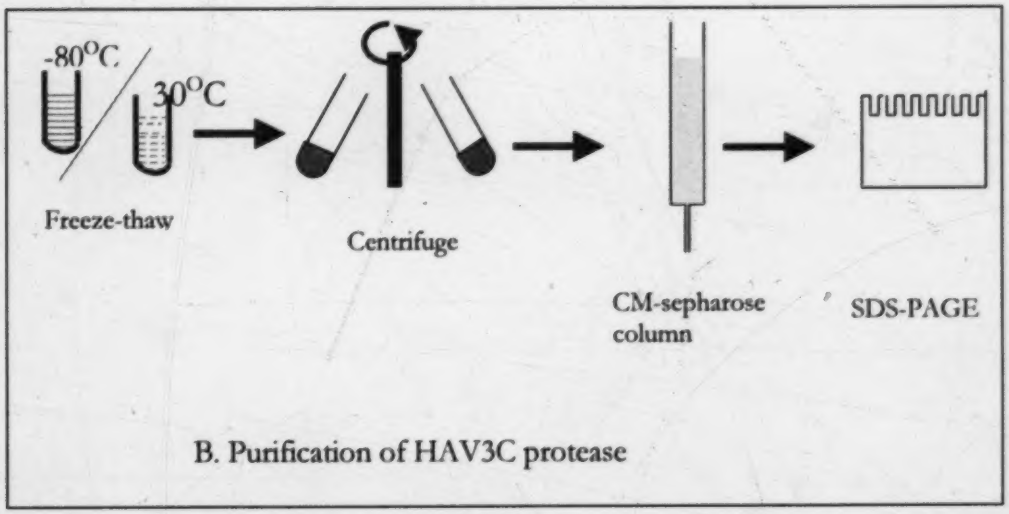


Figure 2.2. HAV 3C protease expression (A) and purification (B) protocol

(see Table 2.1) were inoculated with 10 mL of the overnight culture. The initial $OD_{600} \approx 0.1$. The cells were shaken at 300 rpm at 37°C for 4 to 5 hours, until the OD_{600} reached a value of 0.8 whereupon the cells were induced with IPTG (2 mM final concentration) and allowed to grow for 8 hours at 30°C. The cells were harvested by centrifuging at 3000 rpm for 30 minutes. The cell pellet was stored at - 80°C.

2.2.6. Purification of HAV 3C Protease

The cell pellet from a 1-litre culture was removed from the -80°C freezer and warmed to room temperature. The pellet was then resuspended in 100 mL of resuspension buffer containing 80 μg/mL of lysozyme. The cells were agitated into solution and incubated on ice for 30 minutes. A freeze-thaw process was used to break the cell membranes and to release the HAV 3C protease into solution. The cells were placed onto dry ice for 20 minutes and then thawed in a 30°C water bath for another 30 minutes, a process that was repeated three times. The suspension became progressively more viscous after successive freeze-thaws. After the final thaw, the cell mixture was centrifuged at 15,000 rpm for 1 hour. This step pelleted the cell membranes and organelles, leaving a protein-rich supernatant. After centrifugation, the supernatant was poured into centrifuge tubes and 0.5 mL of 10% polyethylenimine was added to precipitate the DNA. The solution was then mixed and later centrifuged at 15,000 rpm for 30 minutes. The clear yellow supernatant was decanted, placed into a tube and kept at 4°C.

A 50-mL CM Sepharose fast flow column was prepared and washed with three column volumes of a wash buffer (1 M NaCl, 20 mM K_2HPO_4, pH 6.5, 0.5 mM EDTA). The column was then washed with one column volume of 70% ethanol followed by two column volumes of sterilized distilled water. Finally, the column was equilibrated with three column volumes of phosphate buffer (20 mM K_2HPO_4, pH 6.5, 2 mM DTT, 0.5 mM EDTA). The HAV 3C protease sample was loaded onto the column at a flow rate of 1 mL/minute using a peristaltic pump. The eluent was monitored at 280 nm using a single path Ultrospec UV/ visible spectrophotometer (Pharmacia Biotech Ltd., Cambridge, England). The column was continuously flushed with phosphate buffer until the UV monitor no longer detected any protein in the eluent (~ 2 hours).

HAV 3C protease was eluted using a continuous gradient from 0 to 0.50 M NaCl in phosphate buffer by using a simple gradient former (Amersham Pharmacia Biotech Inc.) Fractions (8 mL each) were collected and their absorbance at 280 nm measured using an Ultrospec 3000 UV/ visible spectrophotometer (Pharmacia Biotech Ltd., Cambridge, England). The fractions containing HAV 3C protease (as judged by SDS-PAGE using an authentic sample of HAV-3C protease as a reference) were pooled into a 50 mL centrifuge tube. Sodium azide (25 μL of 3% w/v) was added to inhibit bacterial growth, and nitrogen gas was instilled into the tube to limit oxidation. The sealed tube was stored at 4°C. HAV-3C protease is stable for at least one month under these conditions (Watson, 2000).

2.2.7. SDS-PAGE Electrophoresis

1) SDS-PAGE Gel Preparation. The Bio-Rad Miniprotein system was used to prepare and run all SDS gels. A 12% resolving gel was made according to the recipe presented in Table 2.4. The glass plates were assembled using 1 mm spacers and the acrylamide mixture poured into the gap between the glass plates leaving sufficient space (about 1.5 cm) for the stacking gel. Isobutanol (0.5 mL) was overlaid on the resolving gel surface using a Pasteur pipette to remove the acrylamide meniscus. After the acrylamide polymerization was complete (about 30 minutes), excess isobutanol was poured off. The stacking gel was prepared according to the recipe given in Table 2.4 and poured on top of the polymerized resolving gel. A Teflon comb was inserted and left in place until the stacking gel polymerized.

2) Sample Preparation. Whole cell protein samples were standardized for cell growth by pelleting a cell culture with a volume of (1/ OD_{600}) mL. The pellet was then resuspended in 100 µL of loading buffer. For the cell lysate and column fraction samples, 20 µL of each sample was mixed with 20 µL of 2× loading buffer (see Table 2.4). All samples were heated to 100°C for 5 minutes prior to loading.

The SDS-PAGE gel was placed into a mini-protein electrophoresis apparatus (Bio-Rad Laboratories Canada Ltd., Mississauga ON) and ~ 500 mL Tris-glycine electrophoresis buffer (see Table 2.4) added to the top and bottom reservoirs. Protein samples (10 µL) were loaded into the wells, along with 10 µL of a molecular weight standard in a separate well. The gel was run at 100 volt until the

dye front moved into the resolving gel, whereupon the voltage was raised to 200 volt and maintained until the bromophenol blue reached the bottom of the resolving gel (about 1 hour). The gel was removed, placed in a plastic container on a rocking platform and stained with Coomassie Brilliant Blue R250 (see Table 2.4) for 1 hour. The gel was destained by soaking in destaining buffer (See Table 2.4) on a rocking platform for 24 hours.

2.2.8. Analysis of HAV 3C Protease

The HAV 3C band was identified by comparison with a set of SDS-PAGE molecular weight standards and an authentic HAV 3C reference (Figure 2.3). The OD_{280} was measured for all column fractions and was used to estimate the protein concentration according to equations 2.1 and 2.2.

$$\varepsilon = 5690 \times Trp + 1280 \times Tyr + 120 \times Cys \ (M^{-1}cm^{-1}) \qquad \text{(Eq. 2.1)}$$

For HAV 3C protease, MW=23814.3, number of Tyr = 6, number of Trp = 3, number of Cys = 0. The equation therefore reduces to:

$$\text{Concentration (mg/mL)} = OD_{280} / 1.2 \qquad \text{(Eq. 2.2)}$$

2.2.9. NMR Spectroscopy

As judged by the above UV techniques, the protein-rich fractions from the CM Sepharose purification step contained ~ 20 mg HAV 3C protease in ~ 20 mL of buffer (360 mM NaCl, 20 mM KH_2PO_4). The protein was concentrated using an Ultrafree-15 Centrifugal Filter Device (Millipore, United States). The filter was

Table 2.4. Reagents for SDS-PAGE gel.

Solution	Composition	Amount
Resolving Gel	40% acrylamide (29:1)	3.25 mL
	1.5 M Tris pH8.8	2.5 mL
	10% SDS	100 µL
	10% APS	100 µL
	TEMED	4 µL
Stacking Gel	40% acrylamide (29:1)	630 µL
	1 M Tris pH6.8	630 µL
	10% SDS	50 µL
	10% APS	50 µL
	TEMED	5 µL
Protein 1 × Loading Buffer	Tris-HCl (pH 6.8)	50 mM
	Dithiothreitol (DTT)	100 mM
	SDS	2%
	Bromophenol blue	0.1%
	Glycerol	10%
	Store in freezer	
Protein Running Buffer (1 liter) PH 8.4	Glycine	14.4 g
	Tris Base	3.0 g
	SDS	1.0 g
Staining Buffer	95% ethanol	500 mL
	Glacial acetic acid	100 mL
	ddH_2O	400 mL
	Coomassie BB R 250	1.0 g
De-staining Buffer	Ethanol	25%
	acetic acid	10%

first rinsed with 10 mL of phosphate buffer (20 mM KH_2PO_4, 0.5 mM EDTA, pH 6.5) and centrifuged (3000 g) for 10 minutes to remove residual membrane glycerin. The protein sample (15 mL at a time) was loaded and centrifuged at 3000 g, with the volume checked every 10 minutes until it reached 390 μL. The sample was then brought up to 500 μL by adding 10 μL DSS and 100 μL D_2O (99%) to give a final protein concentration of 1.25 mM. The final solution conditions were 360 mM NaCl, 20 mM KH_2PO_4, pH 6.5, 0.25 mM EDTA, H_2O: D_2O ratio 4:1.

All NMR experiments were set up and performed with the assistance of Dr. Monica Li and Dr. Brian Sykes (University of Alberta). ^1H -^{15}N HSQC experiments were carried out at 30°C on a Varian Unity 600 MHz spectrometer equipped with a three-channel 5 mm pulse-field gradient triple resonance probe. The acquisition time was 75 ms and the pulse width was 9.8 μs. The sample was scanned 16 times for each of the 256 increments. The spectral width was 1600 Hz for ^{15}N and 6999.7 Hz for ^1H. Chemical shifts were directly (^1H) and indirectly (^{15}N) referenced to DSS at 0 ppm as described by Wishart et al (1995). Spectra were processed and peak-picked using the VNMR software package (Version 5.1).

2.3. Results

From the SDS-PAGE analysis shown in Figure 2.3, the HAV 3C protease appeared to be expressed in the CT16 cells at levels approaching 10% of the total cell protein. After freeze-thaw extraction, workup and passage through a CM-

cell protein. After freeze-thaw extraction, workup and passage through a CM-Sepharose cation exchange column, HAV 3C protease was obtained at ~ 95% purity as determined by SDS-PAGE. The final yield was determined to be 20 mg / litre of cell culture.

Eight intense peaks were observed in the 2D ^{1}H-^{15}N-HSQC NMR spectrum (Figure 2.4). The total number of phenylalanine residues known to be in HAV 3C protease is exactly eight. This spectrum clearly demonstrates that eight phenylalanine residues were all labelled with ^{15}N and there was neither isotopic scrambling nor dilution. All of the eight phenylalanine peaks have been assigned now by Dr. Wishart's lab (Figure 2.4) (Watson, 2000; Bjorndahl, et al., 2000).

2.4. Discussion

2.4.1. Expression

The expression of HAV 3C protease from the pHAV-3CEX plasmid (Malcolm, et al., 1992) is driven by the *tac* promoter, a hybrid of the *trp-lac* promoter. This promoter is regulated by the *lac* repressor (Figure. 2.1), which can be induced by IPTG. Overall, our results showed that pHAV-3CEX can be successfully transformed into CT16 cells and express HAV 3C protease very well.

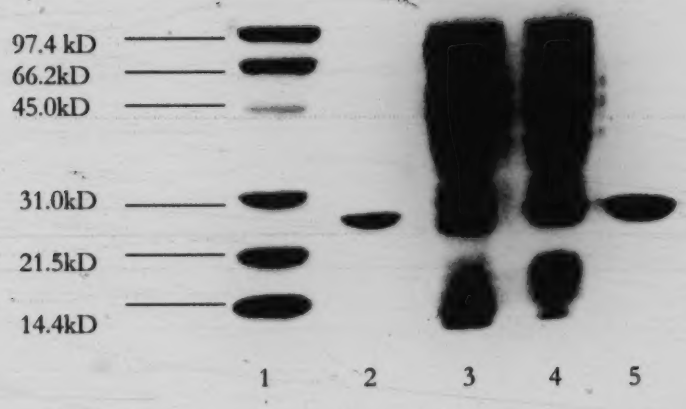


Figure 2.3. SDS-PAGE analysis of HAV 3C protease purification: Lane 1, protein molecular weight standard; Lane 2, authentic HAV 3C reference; Lane 3, whole cell extract without induction; Lane 4, whole cell extract induced for 8 hours; Lane 5, fraction after CM Sepharose purification step.

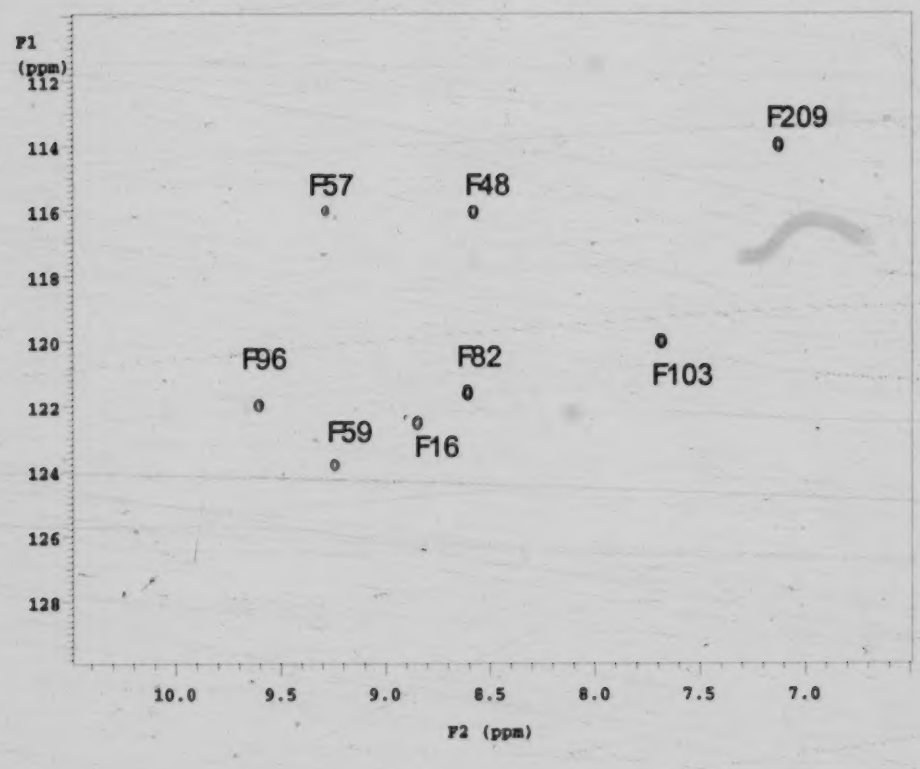


Figure. 2.4. ^{1}H-^{15}N HSQC spectra of ^{15}N-phenylalanine (1 mM) labelled HAV 3C protease collected at 30°C on a Varian Unity 600 spectrometer.

We produced 20 mg of HAV 3C protease from 1 litre of CT16 culture in defined M9 media supplemented with amino acids. This was higher than previously reported yields using the same vector (Malcolm, et al., 1992). Previously reported work using *E. coli* D1210 cells grown in rich media 2 × YT (1.6% bactopeptone, 1.0% yeast extract, 0.5% NaCl) gave average yields of 6 mg of HAV 3C protease per litre of culture (Malcolm, et al., 1992). More recently, Watson (2000) expressed HAV 3C protease in MM294 prototrophic cells and reported yields of ~ 60 mg per litre in M9 media. This extraordinary level of expression reported by Watson may have been due to an adaptive mutation that occurred over the course of multiple generations of growth on minimal media. In terms of induction protocols, Malcolm (1992) described a protocol where cells were induced at an $OD_{650} = 0.4$ and allowed to grow for an additional 6 hours. Subsequent experiments by Watson (2000) indicated that longer induction periods yielded substantially more protein. Therefore, the cells were induced at $OD_{600} = 0.8$ (the mid-log phase of the cell growth curve) and allowed to grow for an additional 8 hours.

The protein yield in CT16 cells was lower than that obtained in MM294 cells. It is possible that the CT16 cells were not fully adapted to the M9 minimal medium, therefore reducing the cell growth and the protease expression level. Because we used Terrific Broth (TB) media instead of minimal media for the overnight culture, the switch from rich media to minimal media could cause a poorer and slower cell growth (Mossakowska, et al., 1997). Actually, it was found

that if minimal media was used instead of TB media for the overnight culture, HAV 3C protease yield in minimal media could be increased by a factor of 2 to 3. (Watson, 2000). Possibly, it is because that cells have a longer period to adapt to minimal media.

2.4.2. Purification

When induced with IPTG, intracellular levels of the HAV 3C protease reached about 10% of total cellular protein as shown in Figure 2.3. A one-step purification protocol via a CM-Sepharose Fast-Flow cation exchanger resulted in a ~ 95% pure HAV 3C protease as judged by SDS-PAGE (Figure. 2.3). The use of a one-step purification protocol partially contributed to the high overall yields, since more

protein would likely be lost through multiple column transfers and column elutions. Earlier reports on HAV 3C protease purification indicated that two chromatographic steps were required to purify the HAV 3C protease, namely a DEAE-Sephadex A25 step followed by a CM-Sepharose column separation (Malcolm et al., 1992). Considering the high cost of ^{15}N labelled amino acids and the need to recover as much protein as possible, we chose to use the fewest purification steps possible. Our single step CM-Sepharose purification strategy proved to work well, so we omitted the DEAE-Sephadex step.

2.4.3. Selective ^{15}N Labelling of Phenylalanine Residues

The biosynthetic pathway of phenylalanine in *E. coli* is illustrated in Figure 2.5. (Waugh, 1996). There are three critical enzymes responsible for the metabolism of phenylalanine: phenylalanine synthetase, aminotransferase A and D. (Umbarger, 1978; Hermann et al., 1983). Phenylalanine synthetase (encoded by genes *pheA*) catalyzes the endogenous synthesis of phenylpyruvate, a direct precursor of phenylalanine. Aminotransferase A and D (encoded by genes *aspC* and *tyrB*, respectively) are involved in the amino group transfer reaction between phenylalanine and glutamic acid. Because of the complexity of its metabolism, selective isotopic labelling with phenylalanine cannot be accomplished by a simple substitution strategy, i.e., by supplying an excess amount of exogenous phenylalanine to inhibit *in vivo* synthesis.

Indeed, earlier efforts aimed at labelling HAV 3C protease with phenylalanine using this simple strategy met with failure, as seen in Figure 2.6. Here the 2D ^{1}H-^{15}N HSQC NMR spectrum shows considerable isotopic "scrambling". A single deficiency in *pheA* cannot prevent ^{15}N scrambling since the labelled amino group can be transferred to glutamic acid, which in turn can be transferred to isoleucine, leucine, valine, aspartic acid, tyrosine, glutamine or proline though aminotransferases A and D. Other researchers have observed a ~ 40% loss of ^{15}N when isotopically enriching T4 lysozyme expression using a *pheA*- phenotype grown on defined media supplemented with ^{15}N phenylalanine (Muchmore et al., 1989). Therefore, both aminotransferases A and D activity must be removed in order to

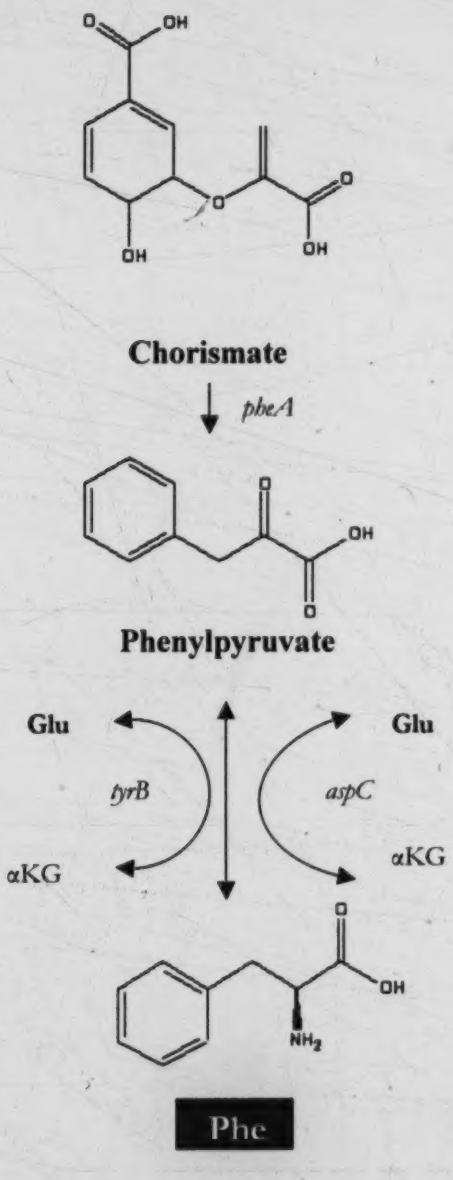


Figure 2.5. Schematic view of phenylalanine biosynthesis in *Escherichia coli*. Pathways are summarized in schematic form, with each arrow representing a discrete biochemical transformation. Genetic acronyms for defective genes that can be used to control amino acid biosynthesis are shown in italics. Abbreviations used are: PHE, phenylalanine; GLU, glutamic acid; α-KG, α-ketoglutarate (Waugh, 1996).

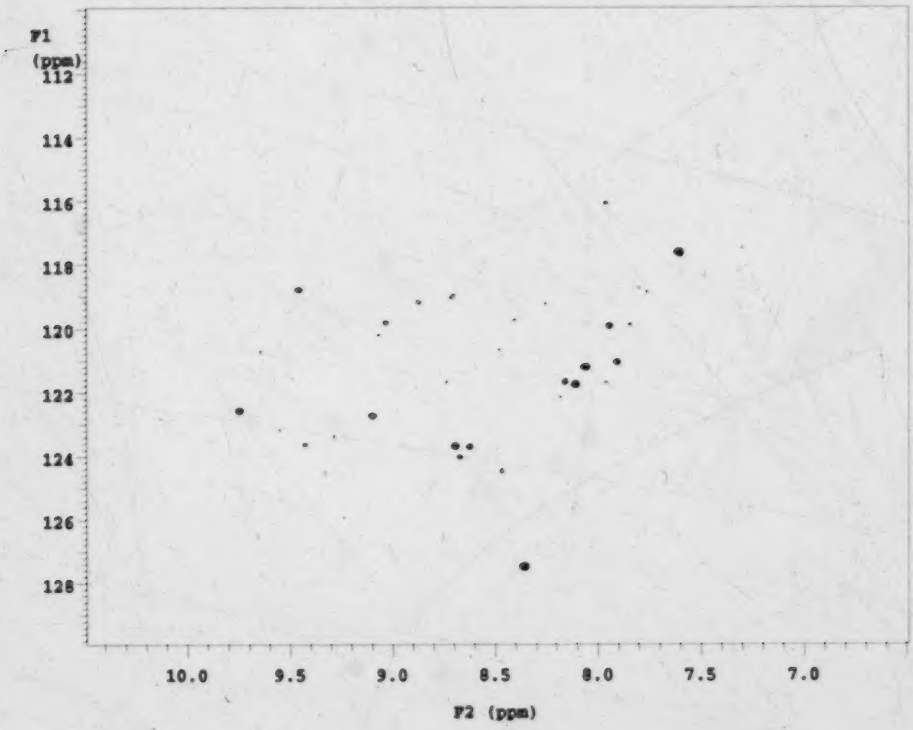


Figure 2.6. ^{1}H-^{15}N-HSQC Spectra of ^{15}N-phenylalanine labelled HAV 3C protease (1mM) at 25°C on a Varian Unity 600 spectrometer. It was obtained using a prototrophic *Escherichia coli* strain (MM294) and by supplying an excess amount of ^{15}N-phenylalanine to the growth medium (Watson, 2000).

isotopically label proteins with phenylalanine. Because of the overlapping specificity of the general aminotransferases, neither the *tyrB* nor the *aspC* mutation alone can confer phenylalanine auxotrophy on an *E. coli* host (Waugh, 1996). Instead, multiple genetic lesions are required. The genotype for CT16 is λDE3; *ilvE*12; *tyrB*507; *aspC*13; *avtA*::Tn5, therefore it is auxotrophic for Phe, Ala, Ile, Leu, Tyr and Val. Although *pheA* is not disabled in this strain, the endogenous synthesis is normally inhibited by the excess amount of phenylalanine added to the media.

2.5 Conclusion

To summarize, we have succeeded in selectively labelling HAV 3C protease with ^{15}N phenylalanine by using the auxotrophic *E. coli* strain CT16. There was neither isotopic dilution nor isotopic "scrambling". All of eight phenylalanine peaks were successful identified and assigned. This work is being used to complete the backbone assignments of HAV 3C protease. This result demonstrates that auxotrophic cell strains can be a useful tool for selective isotopic labelling and that they are indispensable for labelling amino acids that have complex metabolic pathways, such as phenylalanine, glutamic acid, asparagine, serine and leucine. This work opens the door to site-specific labelling attempts for other amino acids in HAV 3C protease and for other picornaviral proteases being studied in our laboratory, such as the human rhinovirus (HRV) 3C protease.

Chapter 3

Purification and isotopic labelling of HRV14 3C

3.1. Introduction

The success of HIV protease inhibitors in the treatment of AIDS has confirmed the feasibility of utilizing viral proteases as drug targets (Mills et al., 1996). As part of a long-term research strategy being developed in Dr. Wishart's lab, we are attempting to extend this approach to the development of inhibitors for other viral pathogens, including HAV (hepatitis A virus) and HRV (human rhinovirus). Similar to HAV, the human rhinovirus encodes a 3C protease that is critical to many respects of its viral life cycle. Just like the HAV 3C protease, the HRV 3C protease is also quite distinct from any known human protease, making it an excellent drug target. Although there are more than 100 serotypes known for HRV, it has been found that all HRV 3C proteases share a good deal of sequence similarity (Matthews et al., 1999). This suggests that a single inhibitor may have good potency for a large number of HRV serotypes. So far HRV14 is the most studied serotype, possibly because of its easy availability. We chose to study the HRV14 3C protease, too, in an effort to develop potential inhibitors specific to this crucial protein.

Although the X-ray crystal structure of HRV14 3C protease has been determined at a resolution of 2.3 Å (Matthews et al., 1994) and unpublished reports indicate that the HRV14 3C protease has been assigned by NMR by

scientists at Agouron Inc., for reasons of commercial confidentiality, neither the coordinates nor the NMR assignments are publicly available. In light of these limitations we have undertaken to assign and determine the solution structure of HRV14 3C protease using multidimensional heteronuclear NMR techniques. However, prior to determining the structure of this protease, an effective expression and purification system must be devised. In this chapter we describe the successful cloning, expression, purification and M9 media conditioning for HRV14 3C protease.

3.1.1. The pET Expression System

We investigated two different expression vectors for HRV14 3C protease: 1) pET3a and 2) pTrcHis2. The pET expression system (Figure 3.1) is a high efficiency system for cloning and expressing recombinant proteins in *E. coli*. It was originally developed by Studier and colleagues (Studier et al., 1986) and is capable of expressing proteins from a variety of prokaryotic and eukaryotic sources. The system consists of the pET plasmid, as well as a *E. coli* strain bearing the DE3 lysogen, such as BL21. The DE3 lysogen encodes the RNA polymerase derived from bacteriophage T7 (T7 RNA polymerase). This gene is under the control of the *lacUV5* promoter, and its expression can be induced by isopropyl-β-D-thiogalactopyranoside (IPTG). T7 RNA polymerase is very selective for T7 promoters and does not recognize the endogenous *E. coli*

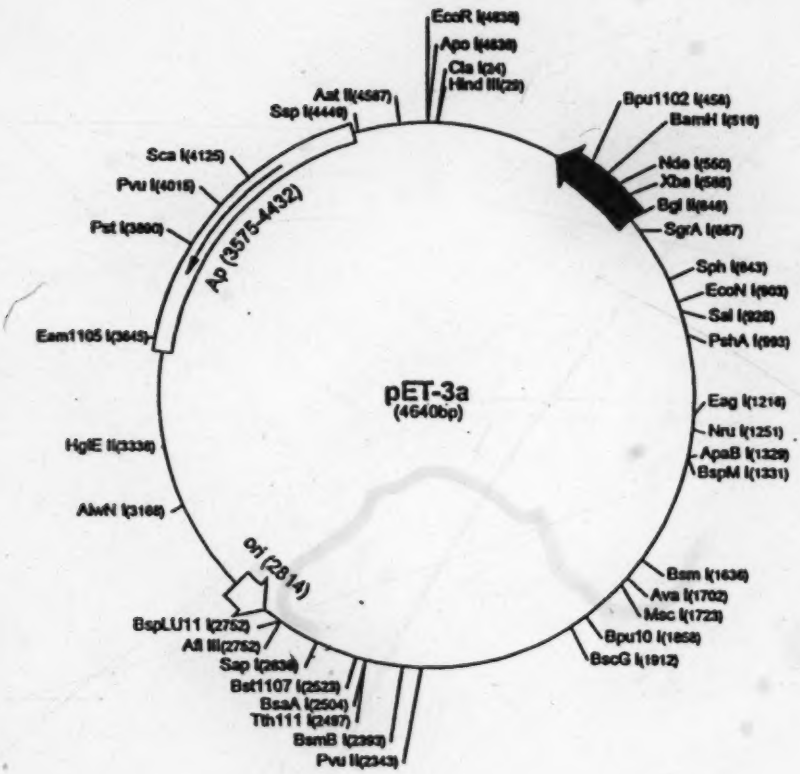


Figure 3.1. pET-3a vector carry an N-terminal T7•Tag sequence and BamHI cloning site. T7 promoter:0 615-631; T7 transcription start: 614; T7•Tag coding sequence: 519-551; T7 terminator: 404-450; pBR322 origin: 2814; *bla* coding sequence: 3575-4432.

promoters. Therefore it will only transcribe genes under the control of the T7 promoter. Consequently the expression of a targeted gene can be extraordinarily high, with the desired product comprising up to 80% of the total cell protein after induction.

The BL21 *E. coli* strain is deficient in the *lon* protease and *ompT* outer membrane protease, both of which can degrade proteins during expression and purification. Some BL21 strains also contain a second plasmid pLysS or pLysE. Both bear the phage T7 lysozyme gene, a natural inhibitor of T7 RNA polymerase. During cell growth and prior to induction, the basal level of T7 RNA polymerase can be inhibited by the endogenously expressed T7 lysozyme. This increases the tolerance of the host cell to toxic target proteins and improves the plasmid retention. T7 lysozyme can also assist in cell lysis by weakening the peptidoglycan layer of the *E. coli* cell wall, which is advantageous in cell lysate preparation.

3.1.2. The pTrcHis2 Expression System

The pTrcHis2 plasmid expression vector (Figure 3.2) is the product of Invitrogen Co. (Carlsbad, CA, USA). It is derived from pUC plasmid. The *trc* promoter combines both *trp* and *lac* promoter components together to make a powerful hybrid promoter that can be induced by IPTG. The *trc* promoter contains the –35 region of the *trp* promoter and the –10 region of the *lac* promoter (Egon et al., 1983). The pTrcHis2 vector also contains a *Lac* repressor

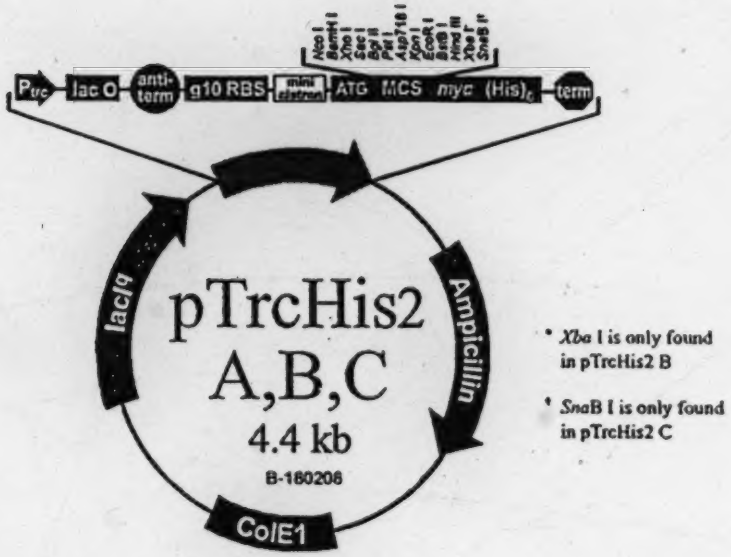


Figure 3.2. pTrcHis2 vector. *trc* promoter region: 190-382; -35 region: 193-198; -10 region: 216-221; *lac* operator: 228-248; *rrn*B antitermination signal: 264-333; gene 10 region: 346-354; Ribosome binding site: 369-373; Expression ATG: 413-415; Multiple cloning site: 411-464; *myc* epitope: 471-503; $(His)_6$ tag: 516-533; *Lac* Repressor ORF: 3285-4367.

(*lacI*^{q}) to regulate expression. Translation will be improved by the bacteriophage T7 gene 10 translation enhancer and a minicistron that provides highly efficient translational restart into the open reading frame of the multiple cloning site. Targeted gene is inserted downstream and in frame with the initiation ATG and a C-terminal fusion peptide. The C terminal peptide contains a *myc* epitope and six histidine residues that function as a Nickel binding site in the expressed protein.

3.1.3. Measuring HRV14 3C Protease Activity

As with any study involving enzyme characterization, the 3C protease activity must be constantly monitored to ensure that the purified protein is both active and in a native conformation. Various methods for measuring 3C protease activity have been reported (Cordingley et al., 1990; Wu et al., 1995; Birch et al., 1995). In 1988 the measurement of 3C protease activity using the intact viral polyprotein or engineered viral polyprotein fragments was proposed (Nicklin et al., 1988). However, although the approach is specific, it is not quantitative. Other methods based on small peptide substrates have been suggested. For instance, it is known that picornavirus 3C proteases expressed in *E. coli* are capable of cleaving short peptides containing sequences of authentic 3C cleavage sites (Cordingley et al., 1989; Knott et al., 1989; Pallai et al., 1989). An assay utilizing these small synthetic peptides as substrates has shown to permit quantification of HRV14 3C protease (Cordingley et al., 1990). In order to design the proper peptide substrate, a series of short synthetic peptides (6-15 amino acids) representing the 2C-3A

authentic protease cleavage site were tested to define the minimal peptide sequences that can be cleaved by HRV14 3C protease (Cordingley et al., 1990). These studies revealed that the cleavage site *in vitro* can only occur at a Gln/Gly scissile bond. Peptides with other authentic cleavage sites (e.g., Gln/Ala or Glu/Gly), can not be processed. Furthermore, a minimum of two amino acids on the carboxyl-terminal side of the scissile bond were required for efficient cleavage. Apparently, proline at the P2' position is also critical as it acts as a helix-destabilizer and strongly influences the secondary structure of the peptide. This work by Cordingley also revealed that the minimum length of the substrate is 6 residues (TLFQ/GP), although longer substrates are favored. A seven-residue sequence (EALFQ/GP) is 10 times more efficient than the minimal substrate and a longer substrate (15 residues) can increase the cleavage efficiency even further. Overall, the best small peptide substrate candidate for HRV14 3C protease is EALFQ/GP.

Picornavirus 3C protease activity assays employing small peptides as substrates are now widely used (Orr et al., 1989; Birch et al., 1995; Wang et al., 1997). Typically HPLC is used to follow the conversion of the substrate to the product during the protease cleavage reaction. This method is quantitative and quite amenable to enzyme kinetic studies (Orr et al., 1989; Long AC, et al., 1989). However, HPLC cannot monitor the peptide cleavage reaction continuously. It is also time-consuming and inconvenient for large numbers of samples. Consequently, two continuous protease assay methods have been

developed (Birch et al., 1995; Wang et al., 1997). One employs a fluorescence method while the other uses a peptide-pNA colorimetric assay. The fluorescence method developed by Birch (1995) utilizes an intramolecularly quenched 9-amino-acid substrate. It has a relatively high background due to the inefficient quenching of the fluorescence donor and quencher. The peptide-pNA colorimetric assay developed by Wang (1997) uses modified peptide EALFQ- p-nitroanilide as the substrate. A chromophoric molecule, p-nitroaniline is placed at the P1' position within the peptide through the formation of an amide bond between the α-carboxyl group of the P1 Gln and the amine group of p-nitroaniline (Figure 3.3). The cleavage of the peptide by the 3C protease will release p-nitroaniline, which differs from its N-acylated derivative pNA substrate by having an absorbance maximum at 405 nm. The extinction coefficient of free p-nitroaniline at this wavelength is 10,360 M^{-1} cm^{-1} at 30°C, whereas that of the peptide substrate-pNA is only 70 M^{-1} cm^{-1}. The large difference in extinction coefficient is the basis of the high sensitivity of this method. Continuously monitoring at 405 nm is accomplished by UV/Vis spectroscopy, which is quick, convenient and accurate. Because of its sensitivity and specificity, we adopted this methodology to detect HRV14 3C activity.

A different approach from the spectroscopic methods described above involves the utilization of a genetic system (Dasmahapatra et al., 1992). This assays involves inserting the 3C protease gene between two domains of the yeast transcriptional activator GAL4. GAL4 contains two separable and functionally

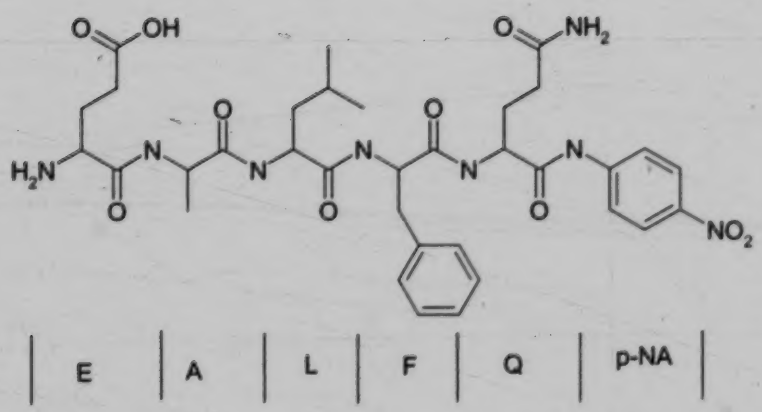


Figure 3.3. Peptide derived HRV14 3C protease substrate EALFQ-pNA

essential domains: the amino-terminal DNA binding domain and the carboxyl-terminal transcriptional activating domain. Dasmahapatra and colleagues constructed a hybrid protein expression plasmid that expresses a chimerical protein containing the autocatalytically active 3C protease (from Coxsackie virus B3) fused in-frame between the two functional domains of the GAL4 protein. The 3C protease can cleave itself out of the fusion protein through an intramolecular cleavage mechanism. This cleavage causes dissociation and inactivation of GAL4, which then fails to activate the transcription of a reporter protein, β-galactosidase. A control chimerical protein which contains a mutated, inactive 3C protease is used to compare activity and monitor the integrity of the assay. The activity of β-galactosidase was determined in quadruplicate, and the average value was expressed in fluorescence units.

3.2 Materials and Methods

3.2.1. Bacterial Strains and Plasmids

The BL21 (DE3) pLysS cells and the pET-3a vector are products of Novagen Inc. (Madison, WI, USA). The pTrcHis2 plasmid was purchased from Invitrogen Co. (Carlsbad, CA, USA). The pET3a-HRV14-3ABC expression system (Figure 3.4) was constructed using PCR cloning techniques by Dr. Ken Ng in the Department of Biochemistry, University of Alberta. The template for the PCR reaction was a cDNA clone of the entire HRV14 genome supplied by

Dr. Wai-Ming Lee (University of Wisconsin). During the expression process, the 3C protease within the 3ABC precursor autocatalytically cleaves itself out of the "triprotein" precursor.

The HRV14 3C protease gene was also subcloned into pTrcHis2 vector (Figure 3.5) by Dr. Dong Luo and Sudeepa Bhattacharyya in Dr. Wishart's lab. The HRV14 3C protease gene was cut from the parent plasmid, pHRV14-3CEX, which was kindly supplied by Dr. Bruce Malcolm (Schering Plough Inc., NJ, USA). It is well known that the protein expression in *E. coli* is often enhanced by changing the nucleotide sequence of the coding region that does not change the amino acid sequence of the expressed protein (Higgins et al., 1999). Expression level could be improved by changing bases G and C to A and T in the first few codons. The possible mechanisms are: 1) some types of codons appear more frequently in the *E. coli* protein genes and *E. coli* cells have more of the corresponding tRNA molecules; 2) changes in the first few nucleotides of gene sequence reduces the secondary structure of mRNA in the initial translation region (Higgins et al., 1999). Based on this hypothesis, several codons at the beginning of the HRV14 3C gene were changed to their degenerative codons, including glycine, proline and threonine. The original sequence at the N terminal is: 5'-ATG GGA CCA AAC ACA GAA-3'. The modified sequence is: 5'-ATG GGC CCG AAC ACC GAA-3'. The modified HRV14 3C protease gene was then

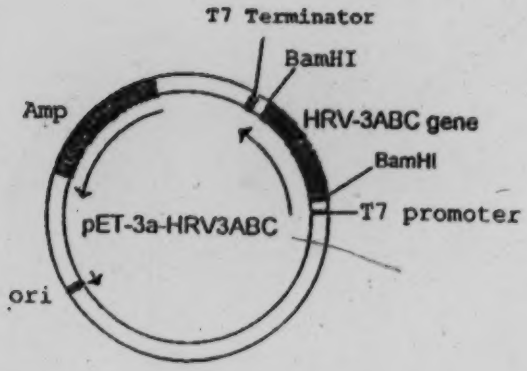


Figure 3.4. HRV14 3C protease expression system in the pET-3a vector. The HRV143ABC gene is controlled under the T7 promoter and inserted between BamHI restriction sites. After expression, the 3C protease will auto-cleave itself from the 3ABC protein to yield the mature protease.

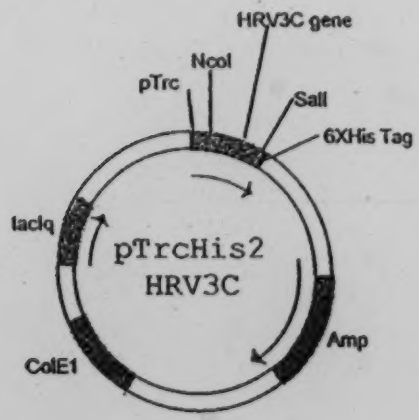


Figure 3.5. HRV14 3C protease expression system in the pTrcHis2 vector. The HRV14 3C gene is under the pTrc promoter. It is inserted between NcoI and SalI restriction sites. There is a stop codon within gene and therefore the Hig-tag will not be translated.

cloned into the pTrcHis2 vector (Figure 3.5) by PCR cloning techniques. The new vector construct was transformed into both BL21(DE3) pLysS and BNN cells and grown in LB media and M9 media.

3.2.2. Expression of HRV14 3C Protease in Luria Broth Media

The pET3a-HRV14-3ABC and pTrcHis2-HRV14-3C vectors were transformed into BL21 (DE3) pLysS cells via an electroporation technique similar to the method described in Chapter 2. A single colony was picked and inoculated into 30 mL of LB media including 100 µg/mL ampicillin. The culture was shaken at 37°C at 300 rpm overnight. The overnight culture was used to inoculate to 1 L of LB media with 100 µg/mL of ampicillin. The cells were shaken at 37°C at 300 rpm until the OD_{600} reached ~ 0.7 (2.5 to 3 hours). IPTG was then added to a final concentration of 1 mM. Cells were induced at 30°C and kept shaking for a further 5 hours. The cells were harvested by centrifugation at 3000 rpm for 40 minutes. The cell pellet was stored at - 80°C. Expression levels were checked by SDS-PAGE.

3.2.3. Adaptation of BL21 (DE3) pLysS Bacteria to M9 Media

After growing in the LB media for 1 hour (225 rpm), freshly transformed BL21(DE3)pLysS cells containing the pET3a-HRV14-3ABC vector were streaked onto a M9 minimal plate containing 100 µg/mL ampicillin. The cells were incubated at 37°C for 2 to 3 days until the colonies grew to 2 mm in

diameter. A single colony was picked, grown in M9 media and the protein expression characterized by SDS-PAGE. An 84% glycerol stock of the M9 adapted cell strain was made and stored at - 80°C.

3.2.4. Purification of HRV14-3C Protease

This protocol was modified from that of Cordingley (1989). The cell pellet from 1 liter of cell culture was thawed from – 80°C to room temperature in a water bath. The cells lysed with a single thawing due to the weakening of the cell wall caused by the presence of T7 lysozyme from the pLysS plasmid. The cell pellet was resuspended in 20 mL of 20 mM imidazole buffer (pH 8.3, 200 mM NaCl) and sonicated on ice with four 20-second bursts. Polyethylenimine (PEM) was added (0.5 mL of a 10% solution) to the cell homogenate and mixed thoroughly for 5 minutes. The cell homogenate was centrifuged at 15,000 rpm at 4°C for 40 minutes to remove nucleic acids. The supernatant (cell lysate) was then dialyzed against 1 liter of 20 mM imidazole buffer (pH 8.3) using a 12,000-14,000 Dalton pore size tubing at 4°C for 4 hours. During dialysis some precipitation was observed. After dialysis the sample was centrifuged at 12,000 rpm for 10 minutes to clarify it.

To complete the purification a 20 mL DEAE-Sepharose Fast Flow column was prepared. The column was pre-equilibrated with five-column volumes of 20 mM imidazole buffer (pH 8.3). The 20 mL protein sample was diluted to 60 mL with the imidazole buffer, then loaded onto the column by gravity (~ 3 mL/min).

The flow-through containing positively charged proteins (mainly HRV14 3C protease) was collected. To precipitate the HRV14 3C protease, the eluent (60 mL) from the DEAE column was mixed with $(NH_4)_2SO_4$ (0.55 g/mL) very slowly and stirred at 4°C for 4 hours. The sample was then centrifuged for 30 minutes (12,000 rpm) to pellet the precipitant. The pellet was stored at - 20°C for later use. The column was regenerated using a four step washing protocol: (1) 5× volume of 6 M urea and 2 M NaCl solution; (2) 5 × volume of 20 mM imidazole buffer; (3) 5× volume of 30% isopropanol; (4) 5× volume of 20 mM imidazole buffer.

A SDS-PAGE gel (12%) was run for the flow-through. The HRV14 3C protease was identified by comparing protein bands (after Coomassie staining) with an authentic HRV14 3C protease reference (supplied by Dr. Ken Ng). The presence of HRV14 3C protease was further confirmed by the colorimetric peptide-pNA 3C protease activity assay. HRV14 3C protease concentration was determined using the BCA protein assay.

3.2.5. Measurement of Protein Concentration

Protein concentrations in the crude cell lysate, the cell lysate after dialysis, and the eluent from DEAE column were measured using a BCA protein assay (Pierce Co., Illinois, USA) (Hinson et al., 1988). The working reagents were prepared in the relative amounts of MA: MB: MC= 50:48:2. For each reaction, 0.5 mL of working reagent, 0.5 mL of ddH_2O and 10 µL of sample were mixed

together, incubated at 60°C for 60 minutes, cooled to room temperature, and their absorbance monitored at 562 nm using a Pharmacia Ultrospec 3000 UV/visible spectrometer. A BSA standard curve was made to allow for the determination of the HRV14 3C protease concentration. The standards were made of BSA at 0.75, 1.5, 3.1, 6.25, 12.5 and 25 μg/mL.

3.2.6. HRV14 3C Protease Continuous Colorimetric Activity Assay

The HRV14 3C protease activity assay was performed at room temperature in a 600-μL reaction mixture containing 25 mM Hepes (pH 7.5), 150 mM NaCl, 1mM EDTA, 6 mM dithiothreitol (DTT), 100 μM substrate EALFQ-pNA, and 90 μL of the 3C protease sample. All the components except the EALFQ-pNA were added to create a reference solution. Hydrolysis of the peptide bond between pNA and glutamine catalyzed by the HRV14 3C protease released yellow-colored free pNA, whose absorbance at 405 nm was continuously monitored by a UV/ visible spectrophotometer.

3.2.7. 1D-^{1}H-NMR Spectroscopy

The protein pellet containing ~ 20 mg HRV14 3C protease was redissolved in the 450 μL H_2O and 50 μL of D_2O. The final pH was 7.4. The NMR experiment was set up and performed by Dr. D. S. Wishart. A 1D ^{1}H NMR spectrum was collected at 25°C on a Varian VXR 500 MHz NMR spectrometer equipped with a

5 mm triple resonance probe. A sweepwidth of 5999.7 Hz was used. A total of 1024 transients were collected. A relaxation delay of 4.0 s was used between transients. The spectra were processed and analyzed using Varian's VNMR software (version 5.1).

3.3. Results

3.3.1. Expression of HRV14 3C Protease

The expressions of HRV14 3C protease in two vectors, pET3a and pTrcHis2, using LB media are shown in Figure 3.6. HRV14 3C protease levels increased with extended induction periods. After being induced with IPTG (1mM) for 5 hours, the pET-3a system obtained HRV14 3C expression with ~ 50% percent of the total cell proteins, while pTrcHis2 got to ~ 20%. Some "leakiness" in expression was evident in both systems, which may have impeded cell growth. SDS-PAGE analysis of the cellular proteins (Figure 3.8) showed that ~ 70% of the HRV14 3C protease was soluble (retained in the supernatant) while a small proportion remained insoluble (likely as inclusion bodies).

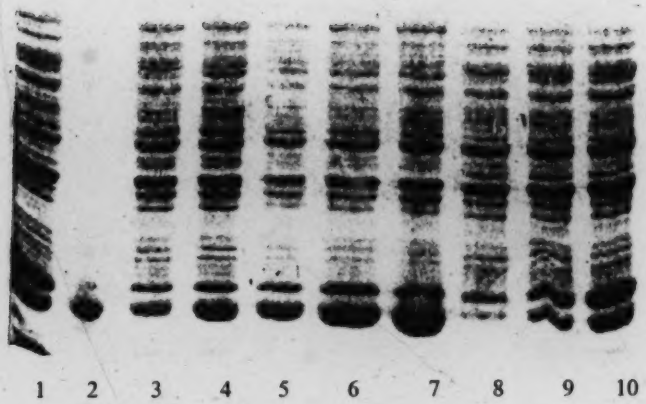


Figure 3.6. SDS-PAGE analysis of HRV14 3C protease expression in *E. coli* using two vectors pET-3a and pTrcHis2. The cells were induced with IPTG for 5 hours at 30°C, after which the whole cell pellets were prepared and analyzed on a SDS-12% polyacrylamide gel. Lanes 3-7 show the expression in the pET-3a. Lanes 1 and 8-10 show the expression in the pTrcHis2 vector. Lane 2: HRV14 3C reference; Lane 3: control cells without induction; Lane 4: after 1 hour induction; Lane 5: after 2 hours induction; Lane 6: after 3 hours induction; Lane 7: after 5 hours induction; Lane 8: control cells without induction; Lane 9: after 1 hour induction; Lane 10: after 2 hours induction; Lane 1: after 5 hours induction.

3.3.2. Adaptation of BL21 (DE3) pLysS Cell to M9 Media

The expression of HRV14 3C in the pET-3a vector from the adapted BL21(DE3) pLysS cell in M9 minimal media is shown in Figure 3.7. After 5 hours of induction using 1 mM IPTG, it appeared that the HRV14 3C expression approached as high as 50% of the total cell proteins, which was almost the same level observed in LB rich media. This suggested that these M9-adapted cells could be suitable for isotopic labelling efforts.

3.3.3. Purification of the HRV14 3C Protease

A significant amount of precipitation was observed during the final dialysis step. This is probably a result of decreased solubility of *E. coli* proteins under low salt conditions. Fortunately, SDS-PAGE revealed that most of precipitated proteins were unwanted cell proteins and only a small amount of HRV14 3C protease co-precipitated (Figure 3.8). Interestingly, the HRV14 3C protease proved to be more soluble at pH 8.3 than at pH 6.8 (Figure 3.9). Consequently this dialysis step served as an effective method for further purifying the HRV14 3C protease. The original intent of the dialysis step was to diminish the salt concentration in the cell lysate and equilibrate it with the DEAE column buffer. However, this step also proved to be advantageous in the purification of HRV14 3C protease. After running the dialysate through the DEAE anion exchanger, the

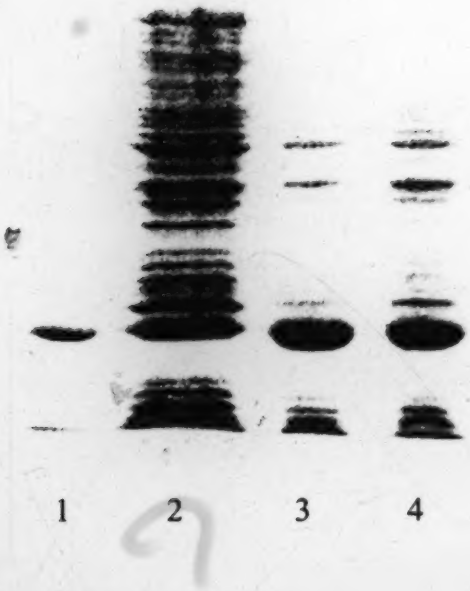


Figure 3.7. Adaptation of BL21 (DE3) *pLysS* cells harboring the HRV14 3C gene to M9 minimal media. Lane 1: HRV14 3C reference; Lane 2: host cell without induction; Lane 3: Expression in LB media after 5 hours induction; Lane 4: the expression of M9-adapted cells in M9 minimal media after 5 hours induction.

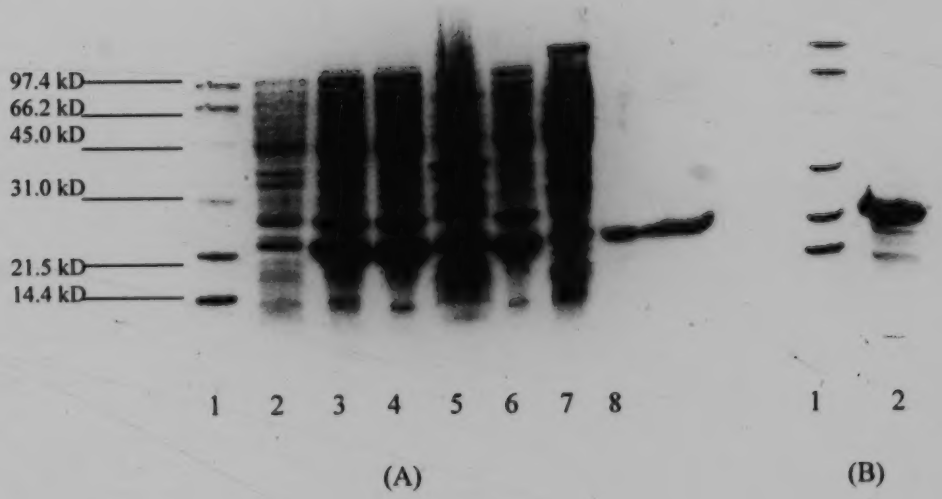
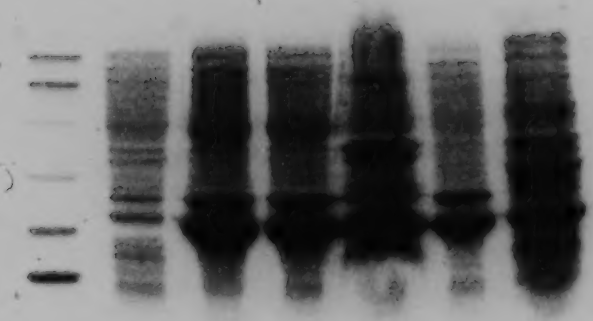


Figure 3.8. SDS-PAGE analysis of HRV14 3C protease purification at pH 8.3 in pET3a system. (A). Lane 1, low range molecular weight standard (composed of phosphorylase B, bovine serum albumin, ovalbumin, carbonic anhydrase, soybean trypsin inhibitor, lysozyme); Lane 2, BL21 (DE3) pLysS host cells (no induction); Lane 3, BL21 (DE3) pLysS whole cell proteins after 5 hours induction; Lane 4, supernatant of cell lysate; Lane 5, insoluble portion of cell homogenate; Lane 6, supernatant after dialysis; Lane 7, precipitated proteins after dialysis; Lane 8, eluent of DEAE-Sepharose column. (B). Purified HRV14 3C protease. Lane 1, low range molecular weight standard; Lane 2, DEAE eluent concentrated 15 times by ammonium sulfate precipitation.



1 2 3 4 5 6 7

Figure 3.9. SDS-PAGE analysis of HRV14 3C protease extract at pH 6.8. Lane 1, low range molecular weight standard; lane 2, uninduced cells (total proteins); lane 3, induced cells (total proteins after 5 hours induction); lane 4, soluble portion of cell lysate; lane 5, insoluble proportion of cell lysate; lane 6, supernatant after dialysis; lane 7, precipitated proteins after dialysis.

HRV14 3C protease was ~ 90% pure (Figure 3.8). The purification profile is listed in Table 3.1.

Table 3.1. Purification profile of HRV14 3C protease.

Steps	Concentration (mg/mL) *	Volume (mL)	Protein Amount (mg)	Purity % †
Cell Lysate	7.8	25	195	40
After Dialysis	2.9	25	72.5	50
DEAE flow through	0.78	60	46.8	90

*It is determined by BCA protein assay.

† It is estimated by SDS-PAGE gel.

The same expression and purification protocol was applied to another HRV14 3C expression system pTrcHis2 that was constructed by Dr. Dong Luo and Sudeepa Bhattacharyya in Dr. Wishart's lab. pTrcHis2-HRV14 3C protease vector was transformed into BNN *E. coli* cells. The result of expression and purification was shown on Figure 3.10. The result showed that this purification protocol could be applied to different HRV14 3C protease expression systems.

3.3.4. HRV14 3C Protease Activity Assay

The result of the HRV14 3C protease activity assay for the eluent from the DEAE-Sepharose column is shown in Figure 3.11. The time dependent change in absorbance at 405 nm had a positive linear relationship within the first 8 minutes of incubation. The initial velocity (V_0) was determined to be 0.020 ΔA/min. This

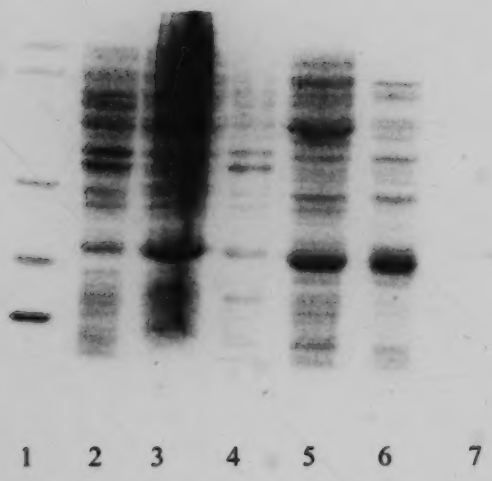


Figure 3.10. SDS-PAGE analysis of HRV14 3C protease purification steps in pTrcHis2 expression system. Lane 1, low range molecular weight standard; Lane 2, BNN host cells (no induction); Lane 3, BNN whole cell proteins after 5 hours induction; Lane 4, insoluble portion of cell homogenate; Lane 5, supernatant of cell lysate; Lane 6, soluble portion after dialysis; Lane 7, eluent of DEAE-Sepharose column.

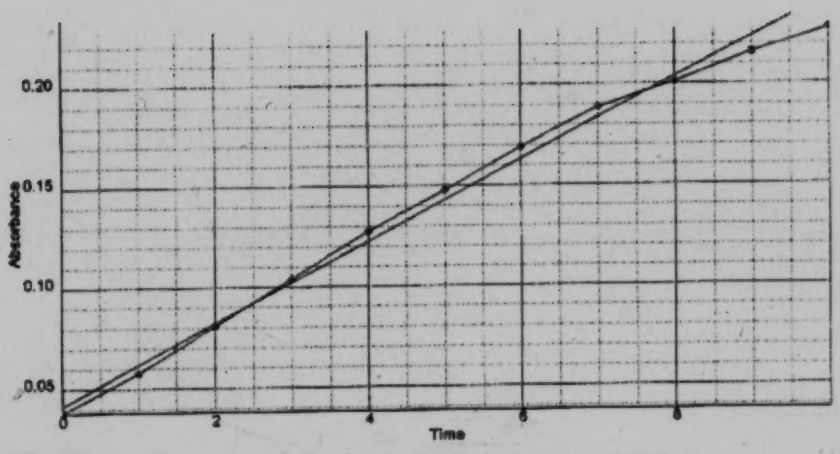


Figure 3.11. HRV14 3C protease colorimetric activity assay for the DEAE eluent: Time dependent curve of pNA peptide cleavage by 3C protease. Substrate concentration = 50 μM. Absorbance at 405 nm was measured at the indicated time points.

activity assay confirmed that the protein in the DEAE flow through was in fact HRV14 3C protease, and that it was active.

3.3.5. 1D ^{1}H NMR Spectroscopy

The ^{1}H NMR spectrum of HRV14 3C protease is shown in Figure 3.12. The spectrum indicated that the protein is properly folded, as seen by the good dispersion of aliphatic (0-4 ppm) and amide shifts (6-9 ppm). The narrow linewidths (10 Hz) are consistent with a monomeric protein of ~ 20 kD. This suggests that the protein should be well behaved for further NMR studies.

3.4. Discussion

3.4.1. Expression of HRV14 3C Protease

The pET-3a vector displayed a higher level of expression for HRV14 3C protease than the pTrcHis2 vector. This perhaps is due to the differences between their promoters, or due to the differences in the protein constructs. The pET3a system contains a polyprotein while the pTrcHis2 vector contains only the HRV3C protease. It may be that the polyprotein protects the HRV14 3C protease against cellular proteases, thereby allows more protein to "survive" the cell for a longer period of time. So we chose to use the pET-3a system to label HRV14 3C protease.

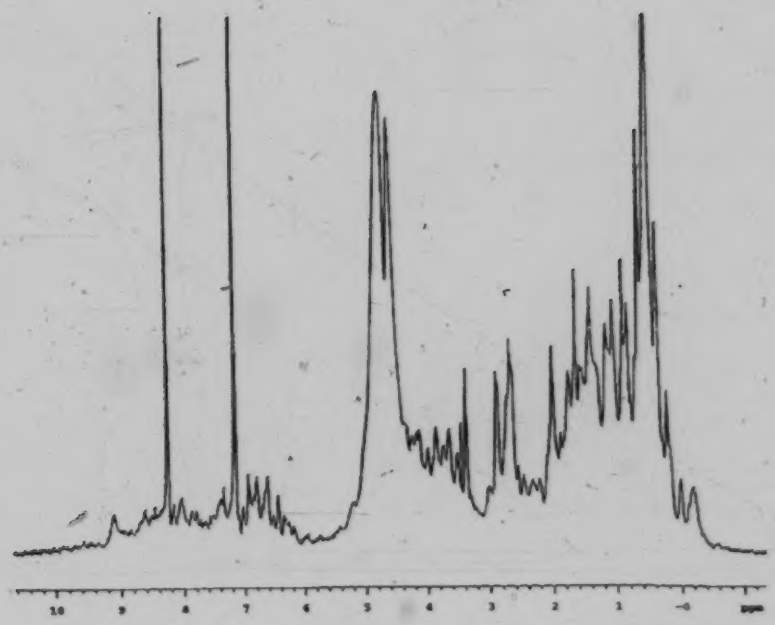


Figure 3.12. 1D ^{1}H NMR spectrum of HRV14 3C protease. The spectrum was collected at 25°C on a Varian VXR 500 MHz NMR spectrometer. The HRV14 3C protease concentration is ~ 0.2 mM.

1) Cell Toxicity of HRV14 3C Protease. The HRV14 3C protease yield is somewhat lower than that of HAV 3C protease (48 mg vs. 100 mg per liter of culture in rich media). It was also observed that the expression levels of HRV14 3C were greatly reduced if the cell colonies were kept at 4°C for more than 1 week. This is an indication of the heterologous protein toxicity or at least a negative selective pressure of 3C protease to *E. coli* cells. In particularly, if mutations that prevent the expression of a toxic heterologous protein occur, then the mutated cells will likely overgrow the desired cell type and greatly reduce the heterologous protein expression level.

To prevent the toxic effects on the cell growth, the basal level of transcription prior to induction ("leakiness") should be suppressed as much as possible. The T7 lysozyme encoded by the pLysS plasmid in BL21 (DE3) pLysS cells is a natural inhibitor of T7 RNA polymerase and will neutralize the basal level of heterologous protein expression. The high level of HRV14 3C protease expression in the pET system suggested that the leakiness in this expression system was well controlled.

2) Inclusion Body Formation. One of the major defects of the *E. coli* expression system is that many recombinant proteins accumulate in the cytoplasm in an insoluble form called inclusion bodies. Inclusion bodies result from accumulation of incompletely folded and misfolded polypeptides (Haase-Pettingell et al., 1988). The intermolecular association of hydrophobic domains during folding

contributes to the formation of inclusion bodies. For proteins with cysteine residues, like HRV14 3C protease, improper formation of disulfide bonds in the reducing environment of the *E. coli* cytoplasm may also contribute to incorrect folding and the formation of inclusion bodies. In our hands (using the pET-3a vector), inclusion bodies accounted for ~ 30% of the total 3C protease produced in the cell. There have been several different reports about the solubility of picornaviral 3C proteases expressed in *E. coli*. Nicklin et al. (1988) expressed poliovirus type 2 3C protease and found that it accounted for only 4% of the total cell protein with most of it being soluble. Ivanoff et al. (1986) expressed poliovirus type 1 3C protease in *E. coli*, but the product was mostly insoluble. Knott et al. (1989) expressed HRV14 3C protease with 50% solubility.

Inclusion body proteins are not active unless they are refolded. However, recovering proteins from inclusion bodies is time consuming and it involves process of isolation, denaturation and refolding (Lilie et al., 1998). The conditions for these steps are often unique to each protein and therefore are difficult to determine. Expressing the protein in a soluble, active form can obviously avoid these problems. A number of strategies have been explored to enhance the yield of soluble, active proteins in *E. coli* as much as possible, these include:

(1) Lowering the cell growth temperature. A number of investigations have suggested a correlation between cell growth temperature and inclusion body formation (Piatak et al., 1988). The interaction between hydrophobic patches on protein surfaces is the primary reason for aggregation and inclusion body

formation. Lower temperatures can decrease the rate of association of folding intermediates, and supply sufficient time for the protein to form native 3-D structure. Therefore, a lower cell growth and induction temperature favors correct folding and the formation of soluble proteins.

(2) Co-expressing with molecular chaperones (Thomas et al., 1996). Chaperones can assist some proteins with overcoming energetically unfavorable protein folding pathways and help to maintain the partially folded proteins in the soluble state.

(3) Fusing proteins with thioredoxin assists in disulfide bond formation (LaVallie et al., 1993). Proteins that are rich in disulfides are often prone to misfolding and incorrect disulfide pairing. This often leads to inclusion body formation. By facilitating disulfide bond formation, these dead-end pathways can be avoided.

(4) Osmolytic stabilizers such as polyols and sugars have been reported to stabilize enzyme activities and secondary/tertiary structure. Bowden et al. (1988) found that the addition of non-metabolizing sugars, such as sucrose, to the medium inhibited the formation of β-lactamase aggregates. Blackwell et al. (1991) also demonstrated that osmotic stress built up by sorbitol is advantageous in expressing large amounts of soluble and active dimethylallyl pyrophosphate 5'-AMP transferase. Kim et al. (1999) reported that when they expressed CGTase in *E. coli*, the addition of mannitol to the medium prevented the formation of aggregates, and raised CGTase's activity by 34 fold. They also found that other

osmolytes including glycerol, erythritol, xylitol, sorbitol and arabitol had the same effect. The mechanisms for this stabilizing effect are not yet clear. A possible explanation is that high osmotic stress strengthens the water lattice structure and decreases the partial molar volume of proteins, therefore diminishing the rate of molecular interactions (Gekko et al., 1981).

3.4.2. Adaptation of BL21 (DE3) pLysS Cell to M9 Minimal Media

M9 minimal media is composed of the basic nutrients required for prokaryotic cell survival. The poor nutritional conditions typically slow bacterial growth and inhibit protein expression because the cells have to synthesize key intermediary metabolites (Mossakowska et al., 1997). There are various methods to facilitate these metabolic changes, but a simple method is to "adapt" the host cells to the minimal media. This can be achieved by swapping the bacteria from rich media to poor media and repeating the swap over several cell growth cycles (Mossakowska et al., 1997).

However, media swapping did not work for our HRV14 3C expression system. This might possibly be due to the toxicity of the HRV14 3C protease. Postulating that young fresh cells would be more adaptable, we performed a transformation on fresh, competent cells and spread the transformed cells on a specifically prepared M9 agar plate. This method turned out to be successful, although it required a longer incubation time (two to three days) to form colonies.

We picked one colony and expressed HRV14 3C at a high level. This indicated that freshly transformed cells were susceptible to adaptation. Another advantage to this approach is that "robust cells", which are adaptable to the minimal media, can be screened at this point.

Another approach of preparing isotopically labelled protein is to grow the host cells in media composed of uniformly ^{15}N, ^{13}C or ^{2}H enriched bacterial, yeast, or algae hydrolysates. The labelled hydrolysates can be produced from organisms grown in defined isotopically labelled minimal media. Some successful examples include: inflammatory C5a protein (Zuiderweg et al., 1989), CMP-KDO synthetase (Fesik et al., 1989) and scFvB80 (Wang Y. J., 1999). This method can often result in higher protein yields than those obtained by growing on minimal media.

3.4.3. Purification of HRV14 3C

A number of different HRV14 3C protease purification strategies have been adopted by different laboratories (Knott et al., 1989; Cordingley et al., 1989; Birch et al., 1995). Since HRV14 3C protease tends to form inclusion bodies in host cells, sometimes with more than half of it existing in the insoluble fraction, there are two different isolation strategies: (1) to purify from the soluble fraction or (2) to purify from the insoluble fraction. Of course, to purify from the soluble fraction is generally more convenient. Knott et al. (1989) purified HRV14 3C from the soluble fraction. In their construct, 50% of the HRV14 3C was soluble.

A cation exchange column (S-Sepharose fast flow) and a gel-permeation column (TSK G3000SW) were combined to get ~ 1 mg of HRV14 3C per liter of culture with 95% purity. Cordingley et al. (1989) used a different protocol to extract HRV14 3C from the soluble fraction. These workers used anion-exchange chromatography (DEAE-Sephacel) at pH 8.3, because most of *E. coli* proteins have isoelectric points below 8.0, they are negatively charged at pH 8.3 and will bind to the DEAE column. On the other hand, proteins with high pIs, such as HRV14 3C protease, are positively charged at this pH and will flow through the column.

The unbound material from this column was concentrated by ammonium sulfate precipitation (0.55 g/mL). They obtained ~ 90% pure HRV14 3C protease at yields approaching 10 mg protein per liter of culture. On the other hand, Birch et al. (1995) chose to work with insoluble fraction to purify their HRV14 3C. The cell granules were solubilized in 7 M deionized urea, then an anion exchange column (Q-Sepharose Fast Flow) and cation exchange column (SP-Sepharose column) were run. After refolding, the protein was further purified by size exclusion chromatography. This protocol yielded 60 mg pure HRV14 3C per liter culture when using a λP_L system, but just 8.5 mg/L when using a pET-3a system.

3.5. Conclusion

HRV14 3C protease was expressed in two different vectors, pET-3a and pTrcHis2 using BL21(DE3)pLysS cells. The expression efficiency in the pET-3a system was found to be higher than that in the pTrcHis2 system. The BL21 (DE3) pLysS host cells were successfully adapted to M9 minimal media by growing freshly transformed cells on an agar plate. The purification process involved a combination of dialysis and anion exchange chromatography using DEAE-Sepharose column. Approximately 46 mg of HRV14 3C protease at ~ 90% purity was obtained from 1 liter of culture. Enzyme activity assays and NMR analysis indicated that the protein was properly folded and fully active. This work should open the way to producing isotopically labelled HRV14 3C protease for further NMR studies.

Chapter 4

Conclusion

Picornaviruses are the causative agents for a number of diseases including poliomyelitis, the common cold, hepatitis A, foot and mouth diseases and cardiomyelitis. Picornaviral 3C proteases appear to be excellent drug targets because of their importance in viral maturation and replication. Their small size (20-24 kD), good solubility and distinct structural characteristics also make them ideally suited for structure aided drug design efforts. In this thesis I have described preliminary results concerning the expression, purification, isotope labelling and spectroscopic characterization of two picornaviral 3C proteases: the HAV 3C protease and the HRV14 3C protease.

Chapter 2 describes how we expressed HAV 3C protease and specifically labelled it with ^{15}N phenylalanine in an auxotrophic *E. coli* cell strain known as CT16. Previous attempts using a prototrophic cell strain (MM294) to specifically label HAV 3C protease with ^{15}N phenylalanine were unsuccessful. This is because phenylalanine metabolism is dependent on two aminotransferases, A and D which will cause ^{15}N transfer to other amino acids. The CT16 is a cell line deficient in both of these aminotransferases, therefore it requires an exogenous source of phenylalanine. No difficulty was found in adapting the CT16 cells to minimal media and the cells harboring the pHAV-3CEX plasmid expressed HAV

3C protease exceedingly well. The resulting protein sample has been used to complete the backbone ^{15}N assignments of HAV 3C protease.

Chapter 3 describes the expression and purification of HRV14 3C protease. BL21(DE3)pLysS cells harboring the pET3a-HRV14-3ABC were successfully adapted to M9 minimal media, though cost more effort than the HAV 3C expression system. This was accomplished by growing the transformed cells on an M9/agar-Amp plate. Our results showed that the freshly transformed cells exhibited better expression levels than glycerol stock cells. Purification of HRV 3C protease was accomplished by devising a simple DEAE filtration strategy. This simple approach resulted in yields of 46 mg/L of 90% pure protein from rich (LB) media. About 30% of the HRV 3C protease forms inclusion bodies.

To complete the ^{15}N labelling of HRV 3C protease should simply require expressing the HRV 3C in the $^{15}NH_4Cl$ substituted M9 minimal media. The host cell BL21(DE3)pLysS has already been adapted and stored in the – 80°C freezer. Once sufficient material has been purified, heteronuclear NMR experiments will be undertaken to complete the assignment and NMR structure determination of HRV14 3C protease. Upon completion of these experiments, inhibitor screening assays will begin using SAR-by-NMR or affinity NMR approaches.

References

Allaire, M., Chernaia, M. M., Malcolm, B. A. & James, M. N. G. (1994) *Nature* (London) 369: 72-76.

Andre, F. E., D'Hondt, E., Delem, A. & Safary, A. (1992) *Vaccine* 10: S160-168.

Bachovchin, W. M. (1985) *Proceedings of the National Academy of Sciences of the United States of America* 82: 7948-7951.

Bax, A. & Weiss, M. A. (1987) *Journal of Magnetic Resonance* 71: 571-575.

Belsham, G. J., McInerney, G. M. & Ross, S. N. (2000) *Journal of Virology* 74: 272-280.

Bergmann, E. M., Cherney, M. M. & Mckendrick, J. (1999) *Virology* 265: 153-163.

Bergmann, E. M., Mosimann, S. C., Chernaia, M. M., Malcolm, B. A. & James, M. N. (1997) *Journal of Virology* 71: 2436-2448.

Birch, G. M., Black, T., Malcolm, S. K. & Lai, M. T., Zimmerman, R. E. & Jaskunas, S. R. (1995) *Protein Expression & Purification* 6:609-618.

Bjorndahl, T. C., Watson, M. S., Slupsky, C. M., Sykes, B. D. & Wishart, D. S. (2000) *Complete backbone chemical shift assignment for the HAV-3C Protease*. PENCE annual general meeting protease symposium, Toronto.

Blackwell, J. R. & Horgan, R. (1991) *FEBS Letters* 295: 10-12.

Bowden, G. A. & Georgiou, G. (1990) *Journal of Biological Chemistry* 265:16760-16766.

Bowen, J. P., Charifson, P. S., Fox, P. C., Kontoyianni, M., Miller, A. B., Schnur, D., Stewart, E. L. & Van Dyke, C. (1993) *Journal of Clinical Pharmacology* 33: 1149-1164.

Brill, B. M., Kati, W. M., Montogomery, D., Karwowski, J. P., Humphrey, P. E., Jackson, M., Clement, J. J., Kadam, S., Chen, R. H. & McAlpine, J. B. (1997) *Journal of antibiotics* 49: 541-546.

Brinkman, K & Kakuda, T. N. (2000) *Current Opinion in Infectious Diseases* 13: 5-11.

Cann, A. J. (1997) *Principle of molecular virology*. Academic Press, San Diego.

Chen, A. & Shapiro, M. J. (1999) *Analytical Chemistry* 71: 669A-675A.

Clark, M. E., Hammerle, T., Wimmer, E. & Dasgupta, A. (1991) *EMBO Journal* 10: 2941-2947.

Cordingley, M. G., Callahan, P. L., Sardana, V. V., Garsky, V. M. & Colonno, R. J. (1990) *Journal of Biological Chemistry* 265: 9062-9065.

Cordingley, M. G., Register, R. B., Callahan, P. L., Garsky, V. M. & Colonno, R. J. (1989) *Journal of Virology* 63: 5037-5045.

Cox, G. A., Fohnson, R. B., Cook, J. A., Wakulchik, M., Johnson. M. G., Villarreal, E. C. & Wang, Q. M. (1999) *Journal of Biological Chemistry* 274: 13211-13216.

Dasmahapatra, B., Didomenico, B., Dwyer, S. M., Sadowski, I. & Schwartz, J. (1992) *Proceedings of the National Academy of Sciences of the United States of America* 89: 4159-4162.

De Febres, O. C., de Petrola, M. C., de Escalona, L. C., Naveda, O., Naveda, M., Estopinan, M., Bordones, G., Zambrano, B., Garcia, A. & Dumas, R. (1999) *Vaccine* 18: 656-664.

Dekker, N., Peters, A. R., Slotboom, A. J., Boelens, R., Kaptein, R. & De H. G. (1991) *Biochemistry* 30: 3135-3146.

Dragovich, P. S., Zhou, R., Webber, S. E., Prins, T. J., Kwok, A. K., Okano, K., Fuhrman, S. A., Zalman, L. S., Maldonado, F. C., Brown, E. L., Meador, J. W. III, Patick, A. K., Ford, C. E., Brothers, M. A., Binford, S. L., Matthews, D. A., Ferre, R. A. & Worland, S. T. (2000) *Bioorganic & Medicinal Chemistry Letters* 10: 45-48.

Edouard, K., Marusyk, R. G., Murphy, F. A. & Van Regenmortel M. H. V. (1990) *Applied Virology Research Volume 2: Virus Variability, Epidemiology, and Control* Plenum Medical Book Company, New York and London. pp.233-259.

Egon, A., Brosius, J., and Ptashne, M. (1983) *Gene* 25:167-178.

Fesik, S. W., Gampe, R. T., Zuiderweg, E. R. P., Kohlbrenner, W. E. & Weigl, D. (1989) *Biochemical & Biophysical Research Communication* 159: 842-847.

Fenner, F., Gibbs, A. & Canberra (1988) *Portraits of Viruses: A History of Virology*. Karger Inc., Switzerland. pp.148.

Gekko, K. & Timasheff, S. N. (1981) *Biochemistry* 20: 4677-4686.

Gosert, R., Dollenmaier, G., Weitz, M. (1997) *Journal of Virology* 71: 3062-3068.

Griffey, R. H. & Redfield, A. G. (1987) *Quarterly Reviews of Biophysics* 19:51-82.

Griffey, R. H., Loomis, R. E., Redfield, A. G. & Dahlquist, F. W. (1985) *Biochemistry* 24: 817-822.

Hajduk, P. J., Dinges, J,, Miknis, G. F., Merlock, M., Middleton, T., Kempf, D. J., Egan, D. A., Walter, K. A., Robins, T. S., Shucker, S. B., Holzman, T. F., & Fesik, S. W. (1997) *Journal of Medicinal Chemistry* 40: 3144-3150.

Hamparian, V. V., Colonno, R. F., Cooney, M. K., Dick, E. C., Gwaltney, J. M., Hughes, J. H., Jordan, W. S., Kapikian, A. Z. et al. (1987) *Virology* 159: 191-192.

Han, N. S. & Tao, B. Y. (1999) *Biotechnology Techniques* 13: 631-635.

Hasse-Pettingell, C. & King, J. (1988) *Journal of Biological Chemistry* 263:4977-4983.

Henry, G. D. & Sykes, B. D. (1990) *Biochemistry* 29: 6303-6313.

Hermann, K. M. & Somerville, R. L. (1983) *Amino Acids: Biosynthesis & Genetic Regulation*. Addison-Wesley Publishing Company, Inc., London, U.K.

Higgins S. J. & Hames, B. D. (1999) *Protein Expression: A Practical Approach*. Oxford University Press Inc., New York. pp.171.

Hinson, D. L. & Webber, R. J. (1988) *Biotechniques* 6:14-19.

Ivanoff, L. A., Towatari, T., Ray, J., Korant, B. D. & Petteway, S. R. (1986) *Proceedings of the National Academy of Sciences of the United States of America* 83: 5392-5396.

Jewell, D. A., Swietnicki, W., Dunn, B. M. & Malcolm, B. A. (1992) *Biochemistry* 31: 7862-7869.

Kapikian, A. Z., Conant, R. M., Hamparian, V. V., Chanock, R. M., Chapple, P. J., Dick, E. C. et al. (1967) *Nature*: 213: 761-763.

Kay, J. & Dunn, B. M. (1990) *Biochimica et Biophysica Acta* 1048: 1-18.

Kay, L. E., Marion, D. & Bax, A. (1989) *Journal of Magnetic Resonance* 84: 72-84.

Kim, M. H., Lee, J. K., Kim, H. K., Sohn, C. B. & Oh, T. K. (1999) *Biotechnology Techniques* 13: 765-770.

Knott, J. A., Orr, D. C., Montgomery, D. S., Sullivan, C. A. & Weston, A. (1989) *European Journal of Biochemistry* 182: 547-555.

LaVallie, E. R., DiBlasio, E. A., Kovacic, S., Grant, K. L., Schendel P. F & McCoy J. M. (1993) *Bio/Technology* 11: 187-193.

Leighton, P. & Lu, P. (1987) *Biochemistry* 26: 7262-7271.

LeMaster, D. M. & Richards, F. M. (1988) *Biochemistry* 27: 142-150.

LeMaster, D. M. & Richards, F. M. (1985) *Biochemistry* 24: 7263-7268.

Levy, G. C. & Lichter, R. L. (1979) *Nitrogen-15 Nuclear Magnetic Resonance Spectroscopy* John Wiley and Sons, Inc., New York.

Lilie, H., Schwarz E. & Rudolph R. (1998) *Current Opinion in Biotechnology* 9: 497-501.

Long, A. C., Orr, D. C., Cameron, J. M., Dunn, B. M. & Kay, J. (1989) *FEBS Letters* 258: 75-78.

Lycke, E. & Norrby, E. (1983) *Textbook of Medical Virology* Butterworths & Co. Ltd. London.

Malcolm, B. A., Chin, S. M., Jewell, D. A., Stratton-Thomas, J. R., Thudium, K. B., Ralstor, R. & Rosenberg, S. (1992) *Biochemistry* 31: 3358-3363.

Malcolm, B. A., Lowe, C., Shechosky, S., McKay, R. T., Yang, C. C., Shah, V. J., Simon, R. J., Vederas, J. C. & Santi, D. V. (1995) *Biochemistry* 34: 8172-8179.

Marion, D., Driscoll, P. C., Kay, L. E., Wingfiled, P. T., Bax, A., Gronenborn, A. M. & Clore, G. M. (1989) *Biochemistry* 28: 6150-6156.

Matthews, D. A., Dragovich, P. S., Webber, S. E., Fuhrman, S. A., Patick, A. D., Zalman, L. S., Hendrickson, T. F., Love, R. A., Prins, T. J., Marakovits, J. T., Zhou, R., Tikhe, J., Ford, C. E., Meador, J. W., Ferre, R. A., Brown, E. L., Binford, S. L., Brothers, M. A., Delisle, D. M. & Worland, S. T. (1999) *Proceedings of the National Academy of Sciences of the United States of America* 96: 11000-11007.

Matthews, D. A., Smith, W. W., Ferre, R. A., Condon, B., Budahazi, G., Sisson, W., Villafranca, J. E., Janson, C. A., McElroy, H. E. & Gribskov, C. L. (1994) *Cell* 77: 761-771.

McIntosh, L. P. & Dahlquist, F. W. (1990) *Quarterly Reviews of Biophysics* 23: 1-38.

Miller, J. H. (1972) *Experiments in Molecular Genetics*. Cold Spring Harbor Press. Cold Spring Harbor, NY, USA. pp 230-235.

Mills, J. S. (1996) *Antiviral Chemical Chemotherapy* 7: 281-293.

Morris, T. S., Frormann, S., Shechosky, S., Lowe, C., Lall, M. S., Gauss-Muller, V., Purcell, R. H., Emerson, S. U., Vederas, J. C. & Malcolm, B. A. (1997) *Bioorganic Medicinal Chemistry* 5(5): 797-807.

Mosimann, S. C., Cherney, M. M., Sia, S., Plotch, S. & James, M. N. G. (1997) *Journal of Molecular Biology* 273: 1032-1047.

Mossakoska, D. E. & Smith, R. A. G. (1997) *Methods in Molecular Biology, Volume 60: Protein NMR Techniques* Humana Press Inc., Totowa, NJ. pp 325-335.

Moy, F. J., Seddon, A. P., Campbell, E. B., Bohlen, P. & Powers, R. (1995) *Journal of Biomolecular NMR* 6: 245-254.

Muchmore, D.C., McIntosh, L. P., Russell, C. B. & Anderson, D. E. & Dahlquist, F. W., (1989) *Methods in Enzymology* 177: 44-73.

Nicklin, M. J. H., Harris, K. S. Pallai, P. V. & Wimmer, E. (1988) *Journal of Virology* 62: 4586-4593.

Opella, S. J., Stewart, P. L. & Valentine, K. G. (1987) *Quarterly Reviews of Biophysics* 19: 7-49.

Orr, D. C., Long, A. C., Kay, J., Dunn, B. M. & Cameron, J. M. (1989) *Journal of General Virology* 70: 2931-2942.

Otting, G., Grutter, R., Leupin, W., Minganti, C., Ganesh, K. N., Sproat, B. S., Gait, M. J. & Wuthrich, K. (1987) *European Journal of Biochemistry* 166: 215-220.

Pallai, P. V., Burkhardt, F., Skoog, M., Schreiner, K., Bax, P., Cohen, K. A., Hansen, G., Palladino, D.E., Harris, K. S. & Nicklin, M. J. (1989) *Journal of Biological Chemistry* 264: 9738-9741.

Pan-Zhou, X-R, Cui, L., Zhou, X-J, Sommadossi, J-P & Darley-Usmar, V. M. (2000) *Antimicrobial Agents & Chemotherapy* 44: 496-503.

Patick, A. K., Binford, S. L., Brothers, M. A., Jackson, R. L., Ford, C. E., Diem, M. D., Maldonado, F., Dragovich, P. S., Zhou, R. & Prins, T. J. (1999) *Antimicrobiology & Agents Chemotherapy* 43: 2444-2450.

Piatak, M., Lane, J. A., Laird, W., Bjorn, M. J., Wang, A. & Williams, M. (1988) *Journal of Biological Chemistry* 263: 4837-4843.

Porter, A. G. (1993) *Journal of Virology* 67: 6917-6921.

Ralstor, R. & Rosenberg, S. (1992) *Biochemistry* 31: 3358-3363.

Renato D. & Harold, S. G. (1988) *Virology* J. B. Lippincott Company. Philadelphia. pp.212-215.

Roeckert R. R. et al. (1996) *Fields Virology*. Raven Publishers, Philadelphia. pp. 609-654.

Salzmann, M., Pervushin, K., Wider, G., Senn, H. & Wuthrich, K. (1998) *Proceedings of the National Academy of Sciences of the United States of America* 95: 13585-13590.

Senn, H., Eugster, A., Otting, G., Suter, F. & Wuthrich, K. (1987) *European Biophysics Journal* 14: 301-306.

Seipelt, J., Guarne, A., Bergmann, E., James, M., Sommergruber, W., Fita, I. & Skern, T. (1999) *Virus Research* 62: 159-168.

Shuker, S. B., Hajduk, P. F., Meadoes, R. P. & Fesik, S. W. (1996) *Science* 274 (5292): 1531-1534.

Skiles, J. W. & McNeil, D. (1990) *Tetrahedron Letters* 31: 7277-7280.

Stanway, G., Joki-Korpela, P. & Hyypia, T. (2000) *Reviews in Medical Virology* 10: 57-69.

Studier, F. W. & Moffatt, B. A. (1986) *Journal of Molecular Biology* 189: 113-130.

Sun, C., Holmgren, A. & Bushweller, J. H. (1997) *Protein Science* 6: 383-390.

Thomas, J. G. & Baneyx, F. (1996) *Molecular Microbiology* 21: 1185-1196.

Timkovich, R. & Cork, M. S. (1982) *Biochemistry* 21: 3794-3797.

Torchia, D. A., Sparks, S. W. & Bax, A. (1988) *Biochemistry* 27: 5135-5141.

Torchia, D. A., Sparks, S. W. & Bax, A. (1989) *Biochemistry* 28: 5509-5524.

Umbarger, H. E. (1978) *Annual Review of Biochemistry* 47: 532-606.

Wang, Q. M. (1999) *Progress in Drug Research* 52: 198-216.

Wang, Q. M., Johnson, R. B., Cox, G. A., Villarreal, E. C. & Loncharich, R. J. (1997) *Analytical Biochemistry* 252: 238-245.

Wang, Y. J. (1999) Ph.D. Thesis: *New Methods for the Characterization of Pharmaceutical Import Proteins.* Faculty of Pharmacy & Pharmaceutical Science, University of Alberta.

Watson, M. S. (2000) Master's Thesis: *Backbone Chemical Shifts Assignments of HAV 3C Protease.* Faculty of Pharmacy & Pharmaceutical Science, University of Alberta.

Waugh, D. S. (1996) *Journal of Biomolecular NMR* 8: 184-192.

Wishart, D. S., Bigam, C. G., Holm, A., Hodges, R. S. & Sykes, B. D. (1995) *Journal of Biomolecular NMR* 5:67-81.

Wlodawer, A. & Vondrasek, J. (1998) *Annual Review of Biophysics and Biomolecular Structure* 27: 249-284.

Wu, Y. & Abeles, R. H. (1995) *Analytical Biochemistry* 229:143-144.

Yalamanchili, P., Datta, U. & Dasgupta, A. (1997) *Journal of Virology* 71: 1220-1226.

Yamasaki, K., Muta, Y. I. Y., Walchili, M., Miyazawa, T., Nishimura, S. & Yokoyama, S. (1992) *Journal of Biomolecular NMR* 2: 71-82.

Zuiderweg, E. R. P & Fesik, S. W. (1989) *Biochemistry* 28: 2387-2391.